

**THE GOOD, THE BAD AND THE CONTENT: BEYOND
NEGATIVITY BIAS IN ONLINE WORD-OF-MOUTH**

A Thesis
Presented to
The Academic Faculty

by

Dezhi Yin

In Partial Fulfillment
of the Requirements for the Degree
Doctor of Philosophy in the
College of Management

Georgia Institute of Technology
August 2012

THE GOOD, THE BAD AND THE CONTENT: BEYOND NEGATIVITY BIAS IN ONLINE WORD-OF-MOUTH

Approved by:

Dr. Han Zhang, Co-Advisor
College of Management
Georgia Institute of Technology

Dr. Samuel Bond, Co-Advisor
College of Management
Georgia Institute of Technology

Dr. Sabyasachi Mitra
College of Management
Georgia Institute of Technology

Dr. Sandra Slaughter
College of Management
Georgia Institute of Technology

Dr. Jack M. Feldman
School of Psychology
Georgia Institute of Technology

Dr. Detmar Straub
College of Business
Georgia State University

Date Approved: May 4, 2012

To my wife Lin,
and my parents Min and Tielin

ACKNOWLEDGEMENTS

Writing this dissertation has been an incredible journey for me. I am sincerely indebted to my committee members, friends, and family for their support and encouragement along the way.

I am especially grateful to my advisors, Han Zhang and Sam Bond, for their mentorship. I thank Han for accepting me into this PhD program, always believing in me, and guiding me through challenges and obstacles throughout the process. I thank Sam for agreeing to supervise me even if I am not in his area, teaching me the ins and outs of experimental design, and standing behind me whenever I need help. They have given me more attention and patience than any one student deserves. Our weekly meetings at room 497 will be my most cherished memory at Georgia Tech. I am very honored to follow in their footsteps.

I am indebted to Michael Cummins, Marius Florin Niculescu, and Lizhen Xu for recruiting their students for my experiments. Without their generosity, I could not have finished this dissertation in time. I also give special thanks to Saby Mitra, Sandra Slaughter, Jack Feldman, Detmar Straub, Eric Overby, Koert Ittersum, Ajay Kohli, Carol Saunders, Jerry Kane, and Xiao Huang, who generously provided constructive suggestions and feedback for some of my papers.

I am very lucky to experience the ups and downs of the graduate program together with Chih-hung Peng, my best friend. We share not only the office but also values and spirits. I will remember our countless conversations over lunches and fun time playing ping pong together. His companionship has made this journey much more wonderful and

exciting. I cherish our enduring friendship, and I look forward to our future collaborations.

I also would like to thank my friends at Georgia Tech, including Lin Jiang, Shanti Agung, Na Wen, Ong-ard Singtokul, Jifeng Luo, Stephen He, Wen Wen, Hui Xu, Yuebing Liu, and my colleagues in the ITM area. I hope them all the best in their academic pursuits.

Finally, I want to thank my wife and parents for their unconditional love and encouragement. I was extremely fortunate and blessed to meet and marry Lin during the PhD program, who has become an inseparable part of my life ever since. I could not have asked for a more supportive companion or a better friend all along the way. I also thank my parents for supporting my decision to study abroad and for always having faith in me. I am proud to have such wonderful parents, and I hope they are proud of me as well.

TABLE OF CONTENTS

	Page
ACKNOWLEDGEMENTS	iv
LIST OF TABLES	x
LIST OF FIGURES	xi
SUMMARY	xii
CHAPTER	
1 INTRODUCTION	1
1.1 Essay 1	2
1.2 Essay 2	4
1.3 Essay 3	6
2 ARE BAD REVIEWS ALWAYS STRONGER THAN GOOD? REVISITING NEGATIVITY BIAS IN ONLINE CONSUMER TRUST	9
2.1 Introduction	9
2.2 Theoretical Framework	12
2.2.1 Negativity Bias and Information Diagnosticity	12
2.2.2 An Attribution-Based Approach	14
2.3 Study 1: Text Reviews	20
2.3.1 Method	20
2.3.2 Results	25
2.3.3 Discussion	27
2.4 Study 2: Profile Ratings	29
2.4.1 Method	30

2.4.2	Results	34
2.4.3	Discussion	36
2.5	Study 3: Trust Recovery	37
2.5.1	Method	38
2.5.2	Results	40
2.5.3	Discussion	42
2.6	General Discussion	43
2.6.1	Theoretical Implications	44
2.6.2	Managerial Implications	46
2.6.3	Limitations	47
2.6.4	Future Research	49
2.7	Conclusion	50
3	SHOULD YOU BE FIRED UP IF YOU WANT TO BE HELPFUL? EFFECTS OF EMOTIONAL AROUSAL IN ONLINE WORD-OF-MOUTH	51
3.1	Introduction	51
3.2	Literature Review and Hypotheses	54
3.2.1	Affect and Emotion	54
3.2.2	Emotional Arousal and the Yerkes-Dodson Law	55
3.3	Study 1: App Store Reviews	57
3.3.1	Data Collection	58
3.3.2	Variables	59
3.3.3	Data Analysis and Results	62
3.3.4	Discussion	66

3.3.5	Lay Theories of Emotional Arousal	67
3.4	Study 2: Experiment	69
3.4.1	Stimulus Materials	70
3.4.2	Procedure	71
3.4.3	Results	73
3.4.4	Discussion	74
3.5	General Discussion	75
3.5.1	Theoretical Implications	76
3.5.2	Practical Implications	77
3.5.3	Limitations and Future Research	78
3.6	Conclusion	79
4	DISCRETE EMOTIONS IN ONLINE SELLER REVIEWS: AN EXPERIMENTAL AND EMPIRICAL EXPLORATION OF REVIEW HELPFULNESS	80
4.1	Introduction	80
4.2	Literature Review and Hypotheses	85
4.2.1	Information Diagnosticity and Affect	85
4.2.2	Theories of Emotion	87
4.2.3	Discrete Emotions and Cognitive Effort	90
4.2.4	Representation of Emotion Concepts	93
4.3	Study 1: Experiment	96
4.3.1	Stimulus Materials	96
4.3.2	Procedure	98
4.3.3	Results	99

4.3.4	Discussion	103
4.4	Study 2: Experiment	105
4.4.1	Procedure	105
4.4.2	Results	106
4.4.3	Discussion	107
4.5	Study 3: Yahoo Merchant Reviews	107
4.5.1	Data Collection	108
4.5.2	Variables	109
4.5.3	Data Analysis and Results	114
4.5.4	Discussion	116
4.6	General Discussion	117
4.6.1	Theoretical Implications	117
4.6.2	Practical Implications	119
4.6.3	Limitations and Future Research	121
4.7	Conclusion	123
APPENDIX A: FOR CHAPTER 2: DEPENDENT VARIABLES AND COVARIATES		124
APPENDIX B: FOR CHAPTER 3: VARIABLES MEASURED IN THE EXPERIMENTS		126
APPENDIX C: FOR CHAPTER 4: VARIABLES MEASURED IN THE EXPERIMENTS		128
REFERENCES		130

LIST OF TABLES

	Page
Table 1: Baseline Reviews	21
Table 2: Treatment Reviews	23
Table 3: Items Used in Store Profiles	30
Table 4: Ratings Used in Store Profiles	33
Table 5: Items Used in Store Profiles	38
Table 6: Descriptive Statistics for Final Review Pool	59
Table 7: Variable Correlations for Final Review Pool (N = 414,336)	60
Table 8: Variables and Operationalizations	62
Table 9: Tobit Analysis Results for Final Review Pool	64
Table 10: Treatment Reviews for App Experiment	72
Table 11: Sample Emotions Characterized by Valence x Uncertainty	91
Table 12: Review Stimuli Used in Experiment	97
Table 13: Descriptive Statistics for Final Review Pool (N = 7,322)	110
Table 14: Variable Correlations for Final Review Pool (N = 7,322)	110
Table 15: Examples of Emotional Review	112
Table 16: Variables and Operationalizations	113
Table 17: Tobit Analysis Results for Final Review Pool	115

LIST OF FIGURES

	Page
Figure 1: Asymmetric Negativity Biases in Trust	27
Figure 2: Screenshot of Profile's Competence Items	32
Figure 3: Asymmetric Negativity Biases in Trust and Purchasing Intention	36
Figure 4: Screenshot of Profile's Items for Participants in Integrity Condition	39
Figure 5: Trust and Purchasing Intention by Profile Dimension	41
Figure 6: Screenshot of Two App Reviews for "Fruit Ninja"	58
Figure 7: An Approximate Plot of the Non-linear Relationship Between Arousal and Review Helpfulness	66
Figure 8: Description Screenshot of the "Fish Pond in Pocket" App	71
Figure 9: Theoretical Framework	95
Figure 10: Perceived Helpfulness of Seller Reviews Across Emotion Conditions	101
Figure 11: Perceived Cognitive Effort of Reviewers Across Emotion Conditions	102
Figure 12: Screenshot of a Yahoo Retailer Review	108

SUMMARY

My dissertation aims to contribute to a more comprehensive understanding of how consumers make sense of online word-of-mouth. Empirical studies have repeatedly confirmed a negativity bias by which “bad” reviews from prior customers have stronger impact than “good” reviews. Nevertheless, few studies have looked beyond this valence-based approach or examined the rich textual content readily available in online reviews. Given the unavoidable nature of negative reviews, online retailers with limited time and resources need more nuanced recommendations for dealing with them. To address these gaps, I explore the boundary conditions of negativity bias in consumer trust formation, and I examine the impact of emotional arousal and discrete emotions on the perceived diagnosticity of review information. Each essay in my dissertation probes beyond the effect of rating valence and explores the role of textual content.

In the first essay, I explore negativity bias among online consumers evaluating peer information about potential sellers. Drawing on research from impression formation and attribution theory, I propose that although online consumers should be biased towards negative content when evaluating information concerning sellers’ integrity, the bias should be reduced for information concerning sellers’ competence. Moreover, decrements in trust and purchasing intention caused by a negative review should be more resistant to change if the review is related to integrity than if it is related to competence. In three experiments, participants were provided text reviews (Study 1) or ratings (Study 2 and 3) of sellers in typical online settings, and then asked about their trust toward the

sellers. Results supported the proposed hypotheses, suggesting that the universality of negativity bias in a seller review setting has been exaggerated.

In the second essay, I examine the impact of emotional arousal on the perceived helpfulness of text reviews. Building on dimensional theories of emotion and the Yerkes-Dodson law, I propose an inverse U-shaped relationship by which the arousal conveyed in a text review will be associated by readers with lower perceived helpfulness only beyond an optimal level. Furthermore, I propose that the detrimental effect of arousal is present for negative reviews even when objective review content is controlled for, and that perceptions of reviewer rationality underlie this effect. To test these hypotheses, two studies were conducted in the context of Apple's mobile application market. In Study 1, I collected actual review data from Apple's App Store, coded those reviews for arousal using text analysis tools, and examined the non-linear relationship between arousal and review helpfulness. In Study 2, I experimentally manipulated the emotional arousal of reviews at moderate to high levels while holding objective content constant, and I measured the process variable of perceived rationality. Results were largely consistent with the hypotheses. This essay reveals the necessity of considering emotional arousal when evaluating review helpfulness, and the results carry important practical implications.

In the third essay, I explore effects of the emotions embedded in a seller review on its perceived helpfulness to readers. Drawing on frameworks in the emotion and cognitive processing literatures, I propose that over and above the well-known negativity bias, the impact of discrete emotions in a review will vary, and that one source of this variance is perceptions of reviewers' cognitive effort. I focus on the roles of two distinct,

negative emotions common to seller reviews: anxiety and anger. In Studies 1 and 2, experimental methods were utilized to identify and explain the differential impact of anxiety and anger in terms of perceived reviewer effort. In Study 3, actual seller reviews from Yahoo! Shopping websites were collected to examine the relationship between emotional review content and helpfulness ratings. These findings demonstrate the importance of discriminating between discrete emotions in online word-of-mouth, and they have important repercussions for consumers and online retailers.

CHAPTER 1

INTRODUCTION

The term ‘negativity bias’ captures a well-accepted, interdisciplinary phenomenon by which “bad things will produce larger, more consistent, more multifaceted or more lasting effects than good things” (Baumeister et al. 2001, p. 325). The tendency to overweigh negative information has not only been established as a general principle of information processing and impression formation (Kahneman and Tversky 1979; Rozin and Royzman 2001), but has also been repeatedly confirmed by empirical studies in online word-of-mouth (Ba and Pavlou 2002; Chevalier and Mayzlin 2006; Pavlou and Dimoka 2006; Zhu and Zhang 2010)

Existing investigations of negativity bias focus largely on concrete and observable variables, such as ratings and reviewer characteristics (e.g., Chevalier and Mayzlin 2006; Forman et al. 2008; Mudambi and Schuff 2010). In addition to these variables, however, consumers also consult the textual content of reviews when making their decisions (Pavlou and Dimoka 2006). The extent of negativity bias indicated in prior literature may therefore depend on aspects of review content. In particular, reviewers often express their emotions freely in review text (Hennig-Thurau et al. 2004), and readers are likely to perceive that emotional content as useful for understanding the reviews and making better decisions (Cao et al. 2011; Kuan et al. 2011). Thus, the emotions embedded in online reviews may have a crucial impact on consumers’ information processing and impression formation.

In summary, relatively few studies of negativity bias have looked beyond ratings or examined the rich textual content readily available in online reviews. In this dissertation, I ask the following question: beyond the effect of valence, how does the content of online reviews influence the way consumers make sense of online word-of-mouth? This question is examined in three essays.

1.1 Essay 1

The first essay demonstrates ‘negativity dominance’ in electronic commerce and explores its operation in the formation of consumer trust. Various criteria have been utilized to identify the existence of negativity bias. One such criterion is ‘negative potency,’ which occurs when negative events are more subjectively impactful than objectively equivalent positive events. In contrast, ‘negativity dominance,’ which occurs when combinations of negative and positive events yield evaluations that are lower than the sum of individual subjective values of those events, is “the most robust and most common exemplification of negativity bias” (Rozin and Royzman 2001, p. 299). However, no research in online word-of-mouth has provided direct evidence for negativity dominance.

More importantly, although empirical studies have provided evidence for a stronger impact of negative ratings than positive ratings on price premiums and sales (Ba and Pavlou 2002; Chevalier and Mayzlin 2006; Pavlou and Dimoka 2006; Zhu and Zhang 2010), very little effort has been made to investigate its boundary conditions. Drawing from research in impression formation and causal attribution, I propose that both the likelihood of negativity bias and resistance to change after a trust violation will

depend on the domain of information discussed in a review. In particular, I examine two fundamental domains — integrity and competence — and argue that negativity bias is more prominent for information regarding sellers' integrity than for information regarding their competence.

To test these hypotheses, I conducted three experimental studies in which participants received feedback about potential sellers in a typical online word-of-mouth setting. In the first two studies, the domain of seller behavior (competence vs. integrity) was directly manipulated through either textual content or seller attribute ratings, and negativity bias was examined by simultaneously presenting positive and negative information of equal extremity. Results showed that negativity bias was not universal, but rather depended on the type of seller information being evaluated. In terms of both trust and actual intention to transact with sellers, the bias was more pronounced for information related to integrity than for information related to competence. Extending these findings, Study 3 explored the dynamic 'trust recovery' process in a setting where a negative rating is followed by subsequent positive ratings. Results indicated that impaired trust and purchasing intention can indeed be recovered with a sufficient number of positive ratings, but the recovery process is faster for competence than for integrity.

The first essay extends current understanding of negativity bias in several ways. By considering not only the valence of reviews but also their content, I move beyond demonstrating the existence of the bias, and instead examine whether consumers exhibit different levels of bias depending on the type of information encountered. In doing so, I utilize experimental methods to overcome the limitations of other approaches, and provide evidence for 'negative dominance,' a stronger and more robust version of

negativity bias. These findings should help scholars to re-evaluate the universality of negativity bias in the online environment, and they also carry important repercussions for online firms striving to enhance consumer trust. If not all bad reviews are created equal, then firms should prioritize their online reputation management efforts accordingly.

1.2 Essay 2

This essay continues the investigation of textual review content in Essay 1 by examining the effect of emotions on the perceived helpfulness of text reviews. Online companies and third-party sites often provide voting mechanisms to identify those reviews that are most useful for assisting consumers in their purchase decisions. These voting systems not only bring value to both customers and companies (Cao et al. 2011; Mudambi and Schuff 2010), but also help third-party review providers curate high-quality reviews and increase site ‘stickiness’ (Connors et al. 2011). Although companies are able to utilize voting mechanisms to identify helpful reviews, the accumulation of votes takes time (Zhang and Tran 2010). They may want to identify helpful reviews early on even before votes have accumulated and react accordingly. In addition, customers who wish to leave reviews for the benefit of future shoppers may have little awareness regarding what constitutes a helpful review. Therefore, understanding the antecedents of review helpfulness has important implications for consumers, companies, and review platforms.

Extending earlier work on review helpfulness, I suggest that the emotions embedded in online reviews may have a crucial impact on their perceived helpfulness. Given the importance of emotions in consumer processing, it is surprising that few

scholars have examined the emotional aspects of review content. This essay focuses on emotional arousal, which has received little scholarly attention in online word-of-mouth literature. Reviewers commonly convey their emotions in text reviews (Hennig-Thurau et al. 2004), and this emotional content can be useful for readers to judge review quality and make better decisions (Cao et al. 2011; Kuan et al. 2011). Drawing on dimensional theories of emotion and the cognitive performance literature, I propose an inverse U-shaped relationship between arousal and review helpfulness: the degree of arousal conveyed in a text review will be associated by readers with lower perceived helpfulness only when the arousal is beyond a certain optimal level. Furthermore, I propose that the detrimental effect of arousal is present for negative reviews even when objective review content is controlled for, and that perceptions of reviewer rationality may underlie this effect.

To test these hypotheses, I employed diverse methods and conducted two studies in the context of Apple's app market. Study 1 collected secondary review data from Apple's App Store and coded emotional arousal empirically. Results provided evidence that arousal reduces the perceived helpfulness of online reviews only after a certain level of arousal is achieved. Extending these findings, Study 2 utilized an experimental method and directly manipulated arousal at moderate to high levels. I found that arousal had a damaging effect for negative reviews, and this effect was mediated by perceived rationality.

This essay reveals the necessity of considering emotional factors (in particular emotional arousal) when evaluating review helpfulness. Emotions are extremely prevalent in online word-of-mouth, and I believe it is meaningful and important to

explore their effects. Compared with valence, emotional arousal is a construct rarely studied in literature, but is one that carries essential meaning. My work provides complementary empirical and experimental evidence associating emotional arousal with perceptions of information diagnosticity. The findings could easily be extended to online word-of-mouth settings in other markets, and they offer clear practical implications. For example, firms might manage reviews more effectively by predicting their helpfulness based on their level of expressed emotional arousal. Moreover, third-party review platforms might use these findings to design and implement writing guidelines that improve the quality of posted reviews and increase site ‘stickiness.’

1.3 Essay 3

Extending Essay 2’s investigation of emotional factors, Essay 3 considers the impact of specific emotions in review content and the underlying mechanisms by which they affect perceptions of helpfulness. In contrast to overall ratings, emotions are highly varied and complex, and cannot be reduced to simple ‘positive’ and ‘negative’ (Lerner and Keltner 2000). In particular, numerous types of negative emotion are present in online reviews (anger, anxiety, disgust, etc.). Because these emotions carry distinct interpretations for readers, the effects of even same-valenced emotions may differ in systematic ways (Fontaine et al. 2007; Levenson 1992). Drawing on frameworks in the emotion and cognitive processing literatures, I propose that the effects of specific negative emotions will vary due to underlying perceptions of reviewers’ cognitive effort. In particular, I argue that holding constant the objective information provided, anxiety-

embedded reviews are considered more helpful than anger-embedded reviews, because anxious reviewers are assumed by readers to think more carefully about the review task.

To test these hypotheses, I conducted three studies utilizing both experimental and empirical methods. In Study 1, I conducted a controlled experiment in which anxiety and anger were manipulated directly while controlling for objective review content, and I explored the process by which these affect perceived helpfulness. As predicted, reviews containing content indicative of anxiety were considered more helpful than those containing content indicative of anger, and their differential impact was explained by beliefs regarding the cognitive effort of reviewers. In Study 2, I replicated the results of the first study while using a different manipulation to rule out alternative explanations. In Study 3, I extended the experimental results by examining actual seller reviews from a popular online platform (Yahoo! Shopping), in order to measure the impact of emotional content on ratings of review helpfulness.

This essay has valuable theoretical and practical implications. By revealing that the effects of negative emotions differ from each other in consistent ways, I show that a valence-based approach is not sufficient to explain the roles of highly varied and complex emotional review content on the reader. Rather, above and beyond specific review content, a review writer's emotions can impact perceptions of helpfulness through perceptions of cognitive effort. Although this essay focuses on two specific emotions, the underlying arguments apply to discrete emotions more generally (e.g., sadness, shame, disgust), and my approach offers an efficient means for testing their effects on reader perceptions. More practically, my results imply that 'ranting' about a bad experience may inhibit the ability of reviewers to influence their audience. Hence, angry reviewers might

either avoid explicit expressions or provide more diagnostic information to counteract their effects. Additionally, the results suggest that sellers should pay particular attention to anxiety-embedded reviews when managing their customer communication efforts, as such reviews are likely to be more impactful.

CHAPTER 2

ARE BAD REVIEWS ALWAYS STRONGER THAN GOOD?

REVISITING NEGATIVITY BIAS IN ONLINE CONSUMER TRUST

2.1 Introduction

The present research examines boundary conditions of negativity bias in electronic commerce. Negativity bias represents a well-accepted assumption that “bad things will produce larger, more consistent, more multifaceted or more lasting effects than good things” (Baumeister et al. 2001, p. 325). In fact, the tendency to overweigh negative information has been established as a general principle in the domains of perception, memory, impression formation, and emotional response (Kahneman and Tversky 1979; Rozin and Royzman 2001).

A generalized negativity bias in information processing also applies to online word-of-mouth in business-to-consumer commerce. In situations where consumers have had no prior interaction with an unfamiliar Web vendor, they can consult various forms of information, including peer reviews posted online about the vendor. By providing reliable feedback from previous customers, online reviews help to reduce uncertainty arising from ambiguous online identities and lack of contextual information, alleviate potential customers’ concerns, and boost corresponding trust (Dellarocas 2003).

Empirical studies in e-commerce have provided some evidence for a stronger impact of negative ratings than positive ratings on price premiums and sales (Ba and Pavlou 2002; Chevalier and Mayzlin 2006; Pavlou and Dimoka 2006; Zhu and Zhang 2010).

Despite abundant empirical evidence indicating the presence of negativity bias, the universally accepted demonstration is ‘negative potency,’ which occurs when negative events are more subjectively impactful than equivalent positive event. In contrast, ‘negativity dominance’ occurs when combinations of negative and positive events yield evaluations that are lower than the sum of individual subjective values of those events. Examining negativity dominance is theoretically important, because it is “the most robust and most common exemplification of negativity bias” (Rozin and Royzman 2001, p. 299). Surprisingly, however, no research in online word-of-mouth, to our knowledge, provided direct evidence for negativity dominance.

In addition, very few studies have questioned the universality of negativity bias, and very little effort has been made to investigate its boundary conditions in consumer behavior (except for some experimental works; e.g., Ahluwalia 2002). Existing investigations of negativity bias focus exclusively on numerical ratings, which are easily observable (e.g., Chevalier and Mayzlin 2006). Notably, however, consumers also consult textual content of reviews in addition to ratings when making their decisions (Pavlou and Dimoka 2006). The extent of negativity bias indicated in prior literature may depend on review content. Furthermore, biases in consumer trust formation remain largely unexplored. Trust is generally defined as “the willingness of a party to be vulnerable to the actions of another ” (Mayer et al. 1995, p. 712). Measurements of trust are generally unavailable in studies using secondary data, and the field study methods using survey data collection techniques commonly utilized in the trust literature are not ideal for capturing the nuances of potential biases. Therefore, an urgent need exists to

probe beyond prior main effects of negativity bias and examine its possible moderators when online consumers develop trust in e-commerce.

Finally, although trust is dynamic in nature, few studies have explored this dynamic process, and almost none address the important question of how shoppers respond to evidence of a trust violation. Sellers attempting to deal with online feedback are faced with an increasing number of reviews that are updated constantly. Given the inevitability of negative reviews, an important question is the extent to which any resulting harm to consumer trust can be overcome by subsequent positive information. Unfortunately, the notion of general negativity bias does little to guide sellers in their ongoing coping efforts. If the effects of exposure to a single, negative review are easily counteracted by exposure to other, positive reviews, then the existence of a few negative reviews is probably not a major concern. On the other hand, if the detrimental consequences of negative reviews are resistant to change, sellers may need to take remedial and more drastic action.

Addressing these gaps, we utilize an experimental approach to answer the following questions: Is negativity bias universal in the formation of trust by consumers exposed to online reviews? More specifically, are negative reviews always more influential than positive ones when both are presented simultaneously? Are the effects of negative reviews equally resistant to change? Drawing from research in impression formation and causal attribution, we propose that both the likelihood of negativity bias and resistance to change after a trust violation will depend on the domain of information discussed in a review. In particular, we examine two fundamental domains — integrity and competence — and argue that negativity bias is more prominent for information

regarding sellers' integrity than information regarding their competence. Our hypotheses are tested directly in three experimental studies that capture or reveal potential biases in realistic consumer settings.

2.2 Theoretical Framework

2.2.1 Negativity Bias and Information Diagnosticity

Trust toward a seller is most often developed through repeated interactions. When no such interactions have occurred, however, trust formation depends critically on information that a buyer acquires about the seller. In the online environment, an especially valuable source of seller information is provided by third-party customer feedback mechanisms (e.g., Bizrate.com, ResellerRatings.com, angieslist.com), which provide a large collection of independent seller reviews. Consumers who utilize these feedback mechanisms are faced with the task of processing and integrating a mix of positive and negative opinions expressed by prior customers.

Importantly, the processing of positive and negative peer information may not occur in a symmetric manner. Abundant evidence supports the existence of a generalized negativity bias in information processing and other domains (Baumeister et al. 2001). In particular, according to Rozin and Royzman (2001), negativity bias can be exemplified in various ways, including negative potency and negativity dominance. Negative potency claims that negative information is more potent than positive information of equal objective magnitude, as described in prospect theory and loss aversion literature (Kahneman and Tversky 1979). Numerous empirical studies in online word-of-mouth provided evidence for the stronger impact of negative ratings on review helpfulness and

sales (Ba and Pavlou 2002; Chevalier and Mayzlin 2006; Mudambi and Schuff 2010; Pavlou and Dimoka 2006; Zhu and Zhang 2010), and few exceptions were observed. In contrast, negativity dominance occurs when the holistic perception of positive and negative information is more negative than the sum of the subjective values of individual pieces of information. Negativity dominance represents the most robust demonstration of the positive-negative asymmetry. However, negativity dominance can only be manifested when positive and negative information are presented together, and thus was rarely examined in literature.

Several theories have been offered to account for negativity bias (see Skowronski and Carlston 1989 for a review). One prominent account, the information diagnosticity perspective, argues that perceivers utilize incoming information about a target to classify the target into one or more behavioral domains (Skowronski and Carlston 1987). The weight attached to a piece of information is dependent on its diagnosticity, defined as “the degree to which one piece of information implies or determines one’s response to a given question or other circumstance requiring a judgment or behavior” (Feldman 1999, p. 48). In other words, a piece of information is diagnostic if it is informative and useful for judgment. The central tenet of this approach is that in general, negative information is more diagnostic than positive information. A major reason is that negative information is generally less common and distinctive. Novel cues can not only increase a perceiver's ability to distinguish among diverse categories of the evaluated target, but also attract more attention and elaboration (Fiske 1980). Therefore, the rarity and novelty of negative information leads it to be perceived as more informative. Another probable reason is that negative information is more certain, carrying a narrower range of potential implications

(Birnbaum 1974; Wyer 1973). Cues that are less ambiguous in terms of potential implications enable a perceiver to make impression judgments more confidently, enhancing its diagnosticity. As a result, less ambiguity of negative information can lead to negativity bias.

For consumers encountering online word-of-mouth about a seller, the information diagnosticity perspective suggests that positive reviews are less diagnostic than negative reviews. Prior empirical studies have demonstrated that negative online feedback is much rarer than positive feedback; e.g., in a sample of the eBay reputation system, less than 2% of ratings were negative, and over 98% were positive (Pavlou and Dimoka 2006; Resnick and Zeckhauser 2002). Furthermore, the implications of negative information are typically clear (avoid this seller), while the implications of positive information are more ambiguous. Due to these reasons, a general negativity bias should be expected, where positive reviews are less diagnostic than their negative counterparts and less impactful on trust formation.

2.2.2 An Attribution-Based Approach

Despite the intuitive appeal (and widespread acceptance) of the negativity bias phenomenon, we argue that in its simplest form, the bias fails to capture important elements of the task faced by customers evaluating seller reviews. More specifically, we suggest that the disproportionate weight attached to negative information about a seller will depend heavily on the domain of seller behavior involved. For present purposes, we focus on the domains of competence and integrity, which have received considerable attention in literature on impression formation (Ferrin et al. 2007; Kim et al. 2006; Kim et

al. 2004; Martijn et al. 1992; Skowronski and Carlston 1987). Both competence and integrity are often discussed directly in the content of seller reviews. For example, a negative review implicating a seller's competence is exemplified by the following, which was modified from real reviews in Yahoo! Shopping: "I did not know when the product shipped, was unable to track it or know when to expect delivery. Terrible operation!" In contrast, a negative review implicating a seller's integrity is exemplified by the following: "The flash drive description said that a cord would be included, but it was not in the package. They did not fulfill their responsibility." Our fundamental argument, developed below, is that due to fundamental differences in the competence and integrity domains, negative information pertaining to each dimension will carry different implications regarding future seller behavior. As a result, the domain of behavior described will have non-trivial impact on the magnitude of negativity bias in trust formation and recovery.

Attribution theory provides a rationale for differences in the processing of negative reviews implicating seller competence vs. seller integrity. Taken broadly, attribution theory refers to the cognitive sense-making process by which individuals ascribe an outcome to possible causes of that outcome (Jones and Davis 1966; Kelley 1967). Expectancy violations theory says that attributional inferences are most likely to occur when events are atypical or unexpected (Burgoon 1993). Therefore, positive reviews are generally unlikely to require inference making; in contrast, a consumer exposed to a negative review is likely to make sense of the incident, and in the process make inferences about the traits and intentions of the seller (Ferrin and Dirks 2003). These inferences can change the relative diagnosticity of negative and positive reviews

(as explained later), thus influencing the magnitude of negativity bias. Viewed in terms of Weiner's (1985) popular model, a consumer will evaluate the cause of a negative review along three attributional dimensions: locus of causality, stability, and controllability.

The first of these, locus of causality indicates whether a negative outcome is due to internal forces (e.g., the seller) or external forces (e.g., other parties or situational factors) (Heider 1958). As suggested in Tomlinson and Mayer's (2009) conceptual paper, internal locus is typically assumed for seller reviews (i.e., reviewers blame sellers for the negative outcomes). If an incident was not attributed to the seller by a reviewer, the review may not be rated negative in the first place. Therefore, we focus on the stability and controllability dimensions, and in particular their impact on the perceived diagnosticity of a seller review.

The second attributional dimension, stability, indicates the degree to which the cause is perceived to remain constant or fluctuate over time. Central to our argument is the notion that dishonest behavior by a seller will be perceived as more stable than incompetent behavior (Tomlinson and Mayer 2009). In fact, dishonesty is generally regarded as stable, because people found to be dishonest on one occasion are likely to be deemed dishonest in general (Kim et al. 2004; Skowronski and Carlston 1987). For example, a seller caught exploiting a misleading return policy is likely to commit similar transgressions in the future. In contrast, competence can be developed over time with training and experience (Heneman and Judge 2006), so that incompetent behavior at one point in time may be a poor predictor of later performance. For example, sellers at fault for poor website design and shipping delays may deal with these issues and improve competence over time.

The final attributional dimension, controllability, refers to the extent to which an individual has volitional control over an outcome. As above, we expect that dishonest seller behavior will be perceived as more controllable than incompetent seller behavior. Dishonest behaviors are typically considered controllable, because an individual could have chosen to behave honestly instead (Kim et al. 2004; Simons 2002). Using the example above, sellers with misleading return policy often have full control over their actions but choose to do so intentionally. In contrast, although competence can be developed over time (Mayer et al. 1995), sellers may not have much control over competence in the short term. Thus, if a negative review concerns incompetence, the cause may be considered less controllable. Using the example above, a seller is likely to have little control over shipping delays for the time being. To summarize, negative reviews related to seller integrity are considered more causally stable and controllable than those related to seller competence.

The arguments above have important implications for the impact of negative review information on trust, and this domain-specific approach yields a more nuanced set of predictions than that suggested by universal negativity bias. Because integrity (vs. competence) violations are perceived to be more stable and controllable, they are less ambiguous with narrower potential implications. As a result, negative reviews concerning seller competence will be perceived as less diagnostic than those concerning seller

integrity (see Birnbaum 1974; Wyer 1973)¹. The impact of negative information will depend on the type of seller behavior involved, and in some cases the negativity bias itself may be attenuated (Ahluwalia 2002). In sum, we predict that the disproportionate impact of negative information will depend on the domain involved: consumer trust toward sellers will be more negatively biased when the reviews concern seller integrity than when they concern seller competence. Additionally, Mayer et al.'s (1995) influential framework identifies trust as the most proximal predictor of risk-taking behaviors such as intention to purchase. This logic is consistent with widely cited attitude models (e.g., the Theory of Reasoned Action; Fishbein and Ajzen 1975), which recognize attitude about the target as an important predictor of behavioral intentions. Applied to the present context, we propose the following hypotheses:

Hypothesis 1: The disproportionate impact of negative information on consumers' (a) trust, and (b) intention to purchase is greater for reviews pertaining to seller integrity than reviews pertaining to seller competence.

Our second hypothesis deals with the dynamic process of trust recovery after a violation. Extending the logic above, we argue that the differential weighting of negative competence vs. integrity information will affect the ability of subsequent positive information to 'repair' the damage caused. If negative integrity-related reviews are

¹ Note that our arguments here concern the relative diagnosticity of reviews referring to different information domains, not the extremity of competence and integrity violations. In each of our experiments, the extremity of negative information was held constant.

perceived as more diagnostic than negative competence-related reviews, then the implications of the former will be more resistant to contradiction, even in the face of positive evidence (Skowronski and Carlston 1992). In other words, the number of positive reviews required to overcome the effects of a single negative review on trust will be greater when the review implicates seller integrity than when it implicates seller competence. As before, this difference in trust in the seller should carry through to intention to purchase (Fishbein and Ajzen 1975):

Hypothesis 2: Resistance to change in consumers' (a) trust, and (b) intention to purchase is greater after reading negative reviews concerning seller integrity than after reading negative reviews concerning seller competence.

In order to test these hypotheses, we conducted three experimental studies in which participants received feedback about potential sellers in a typical online word-of-mouth setting. Hypothesis 1 was examined in Study 1, which utilized realistic text reviews. The domain of seller behavior (competence vs. integrity) was directly manipulated by varying the content of the text reviews, and negativity bias was examined by simultaneously presenting positive and negative information of equal extremity. Although extensive pretesting was conducted, the use of text reviews inherently carries certain methodological concerns. Therefore, to replicate Study 1 while addressing these concerns, Study 2 extended the investigation to a different form of feedback: seller ratings. Finally, Study 3 examined our second hypothesis by providing feedback ratings of multiple stores to explore the process of trust recovery.

2.3 Study 1: Text Reviews

In Study 1, participants were presented with a simulated shopping task in which they were exposed to text reviews disclosing information about a potential seller. The seller reviews mimicked those found at third-party rating sites, (e.g., Bizrate.com), which are commonly used as a means of establishing trust in unfamiliar sellers. Participants were exposed to seller reviews in two stages: first, they received a set of carefully pretested ‘baseline’ reviews that conveyed a neutral impression of the seller. Then they received two carefully pretested ‘treatment’ reviews, one positive and one negative (of equal extremity), regarding either seller competence or integrity. If negativity bias occurs in this context, then exposure to the two countervailing reviews will result in diminished consumer trust. More importantly, if our hypotheses are correct, any such bias will be stronger for reviews implicating seller integrity than for reviews implicating seller competence.

2.3.1 Method

2.3.1.1 Stimulus Materials

As a first step in the stimulus construction process, we conducted pretests to develop a set of three baseline reviews that met the following criteria: 1) each contained only general, nonspecific comments; 2) each was relatively neutral in valence; and 3) together, the three reviews provided basic information about the competence and integrity of the seller to enable subjects to form initial trust. In the pretesting, participants (undergraduate students, $N = 46$ in the final round) were first given the same cover story

used in the main experiment (described below), then read three reviews about a seller, and finally rated that seller's competence and integrity (on separate scales from -3 to +3). The reviews were revised after each round of pretesting until average ratings for competence and integrity were not statistically different from 0 ($M_{\text{competence}} = -0.08, p > 0.5$; $M_{\text{integrity}} = -0.10, p > 0.5$). The baseline reviews that resulted are provided in Table 1.

Table 1: Baseline Reviews

#	Review
1	Their website is OK overall but can be a bit of a hassle. It took me a while to find the button when I tried to add a camera to the shopping cart. Check out was relatively straightforward, though.
2	We bought a television that did not work as well as we had hoped. They told us we could return it if we paid for the shipping. It is mentioned in their return policy, but they did not state the return shipping fee clearly upfront.
3	There were a few issues with my order but it went through in the end. Looking back, I would probably make the same decision, but I would only rate them as average.

Next, two pairs of treatment reviews were constructed (one for competence, one for integrity). Pretests were conducted to identify reviews that met the following criteria: 1) each review uniquely discussed aspects of seller competence or seller integrity, 2) each review was extreme (either positive or negative) with regard to the targeted dimension; 3) the reviews in each pair were opposite in valence; and 4) the reviews in each pair were similar in extremity. First, we collected an initial pool of reviews from e-commerce platforms (e.g., eBay.com, Yahoo! Shopping), third-party review sites (e.g., BizRate.com), and prior literature. We targeted reviews that were both informative and unidimensional; in some cases, reviews were divided and further revised until they represented a single dimension. In the pretests that followed, participants (undergraduate

students, $N = 29$ in the final round) read definitions for each dimension and then received three practice reviews to classify (e.g., a practice review implicating both competence and integrity was the following: “Website was easy to use, easy checkout. Everything came as advertised, just like they said. Nothing to complain about.”). After classifying each practice review, participants were given its correct classification and an explanation. Next, participants read and classified our pool of candidate reviews. They also rated the valence and extremity of each review for each dimension chosen: “In your opinion, how incompetent/competent (dishonest/honest) does this review portray the seller to be?” (-4 = “extremely incompetent (dishonest)” and 4 = “extremely competent (honest)”).

We calculated the proportion of participants who correctly classified each candidate review and retained only those reviews that were assigned correctly by over 75% of participants. From the resulting set, we selected two extreme reviews for each category that were opposite in valence but equal in extremity (see Table 2).

Table 2: Treatment Reviews

Category	Review	Extremity
Competence	Our family has ordered numerous items from them. It's easy to do and their order processing is fantastic! Anything you need during the purchase is on their online system. They'll always have my business.	3.2
	They are not capable of packaging at all! Shipping box was sent with no indication that there was a TV in the box or that it may be fragile. As a result, My LCD TV came with obvious shipping damage. They really need to have some basic training on packaging before running a company!	-3.1
Integrity	This vendor is very trustworthy. I needed my camcorder as soon as possible, and their site warns you which items may take longer to ship. Very honest in my opinion.	3.2
	Bait and switch artists. Advertised one thing on their website ... we ordered after making sure it was in stock and what we wanted, but then they tried to send us something else. After I complained, they changed the ad. Beware! Never believe what they say!!	-3.2

Note: The two reviews for each category were not reliably different in magnitude ($p_s > 0.5$).

2.3.1.2 Procedure

Eighty-two undergraduate students from a southern U.S. university participated in the study. The cover story asked participants to assume that they were shopping online for a digital camera. They were told that they had already decided on a model, browsed a number of different online electronics retailers that offered this model, and tentatively selected one particular seller, “ElectronicWorld.com” (a name created for the study).

Before deciding whether to place their order from this seller, participants were asked to visit a third-party review website and read recent customer reviews of ElectronicWorld.com. The first three reviews, seen by all participants, were ‘baseline’

reviews pretested to be neutral in terms of competence and integrity (see above). After reading these baseline reviews, participants rated their trust in the seller. Next, each participant read a pair of ‘treatment’ reviews, one positive and one negative, pertaining to either competence or integrity (see above). The two treatment reviews were presented on the same screen. The order of the positive and negative reviews was counterbalanced to account for the potential recency effect. After reading the two treatment reviews, participants again completed the trust measures. Finally, participants completed individual difference measures (below), responded to an attention check in which they identified the two treatment reviews, and indicated whether they had prior experience with ElectronicWorld.com.

2.3.1.3 Measures

The primary dependent measures assessed trust in the seller. It was measured on a 7-point Likert scale (1 = “strongly disagree” and 7 = “strongly agree”) using three items adapted from Gefen (2000). Because the treatment reviews were pretested to be equal in extremity, they should (on average) induce no change in perceived trust. Therefore, bias was operationalized by subtracting the first evaluation from the second one.

Because both propensity to trust and institution-based trust have been purported to influence trust (McKnight et al. 1998), these were measured at the end of the procedure. Propensity to trust refers to a general willingness to trust others and is considered a stable personality trait. Institution-based trust indicates perceptions of the institutional environment (in this case, internet retailing). Dependent measures and covariates for all studies are presented in Appendix A.

2.3.2 Results

2.3.2.1 Attention Checks

The data were first screened for participants that: 1) failed to correctly identify one or both of the treatment reviews at the conclusion of the study, 2) reported transacting with the fictional store in the past, or 3) took insufficient time reading the treatment reviews (more than two standard deviations below the mean). This process resulted in the exclusion of ten participants, for a final sample size of 72 (38 in the competence condition and 34 in the integrity condition).

2.3.2.2 Reliability and Validity

Examination of the reliability and validity of major constructs in the study revealed that Cronbach's alphas for all constructs were well above 0.70, demonstrating adequate internal consistency (Nunnally 1967). Therefore, the average score of each construct was calculated and used in further analysis.

Next, we conducted a confirmatory factor analysis (CFA) to assess the convergent and discriminant validity of the competence and integrity measures (Bagozzi and Phillips 1982). Allowing perceived competence and perceived integrity to freely correlate with each other, CFA indicated an acceptable fit ($\chi^2 (19) = 25.740, p = 0.138$; RMSEA = 0.070; all item-factor loadings > 0.70), supporting convergent validity. The hypothesized two-factor model fit the data significantly better than a more parsimonious one-factor model ($\chi^2 (1) = 84.015, p < 0.001$). Loadings of items on their corresponding latent factor

were higher than: 1) loadings of other items on this latent factor, and 2) loadings of these items on the other latent factor, indicating adequate discriminant validity.

2.3.2.3 Trust

We examined the extent to which any bias in trust depends on the type of behavior implicated in the reviews. A repeated-measure ANCOVA was performed to examine changes in trust after reading the pair of treatment reviews. Information condition and review order were entered as between-subject factors; propensity to trust and institution-based trust were entered as covariates. In keeping with our hypothesis, the condition \times treatment interaction term was significant ($F(1, 66) = 4.064, p < 0.05$), indicating that declines in trust were impacted differently by reviews regarding seller integrity and reviews regarding seller competence.

Follow-up comparisons revealed that the pattern of the interaction was in line with predictions. For both conditions, a negativity bias was indicated by a reliable drop in trust after reading the treatment reviews. As expected, a reliable negative bias was obtained for the integrity group, as trust was significantly lower after reading the treatment reviews (M_{baseline} vs. $M_{\text{treatment}} = 3.71$ vs. $3.04, p < 0.001$). However, the same was not true in the competence group (M_{baseline} vs. $M_{\text{treatment}} = 3.62$ vs. $3.43, p = 0.25$). Based on a post-hoc analysis, power was sufficient (> 0.8 , $\alpha = 0.05$) to identify a difference of moderate effect size (Cohen's $d = 0.5$), indicating that trust was relatively unaltered by exposure to the treatment reviews. Together, the observed interaction of valence and information domain indicates that trust in the seller was biased towards negative information to a greater extent when that information pertained to seller integrity

than when it pertained to seller competence. The pattern of means for trust in the seller is illustrated in Figure 1.

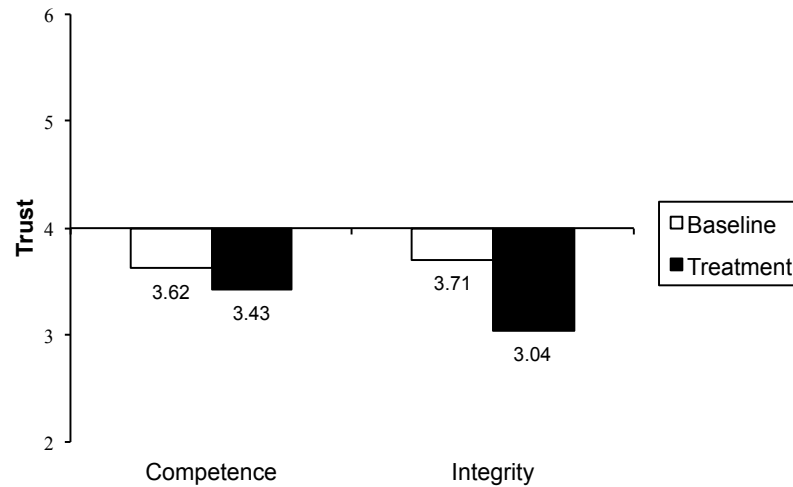


Figure 1: Asymmetric Negativity Biases in Trust

2.3.3 Discussion

Study 1 presented initial evidence that biases in consumer trust perception depend on the type of information encountered. As predicted, analyses of trust revealed a negativity bias that was significantly greater for information regarding seller integrity than for information regarding seller competence; therefore, H1a was supported.

It is important to consider alternative explanations that might be offered to account for the observed asymmetry in negativity bias. First, one might argue that integrity issues are simply more serious than competence issues. However, the major goal of the present paper is to examine negativity dominance, and our hypotheses regard integration of positive and negative information presented simultaneously. Therefore, even if competence and integrity issues differ in importance, this cannot explain why the extent of negative bias was stronger for participants given integrity-related reviews than

for those given competence-related reviews. Moreover, competence-related reviews (both positive and negative) dominate online reputation systems (e.g., Pavlou and Dimoka 2006), suggesting that seller competence is of great concern to online consumers.

Second, participants were asked about their trust in the seller twice before and after they read the two treatment reviews. It is possible that this procedure created a demand effect by exposing the subjects to the dependent measures before they saw the positive and negative reviews. However, competence vs. integrity was manipulated in a between-subjects manner, and the demand effect occurs (if any) in both conditions. Therefore, demand effect cannot explain the differential extent of negativity bias that we observed. Furthermore, measuring trust twice enables us to capture and measure negativity dominance directly. Despite a potential demand effect, we believe this procedure is adequate (if not optimal) given our major goal in this study.

Another concern is that despite extensive pretesting, the two treatment reviews utilized in each pair may have been non-equivalent in aspects other than valence and extremity. Because the study utilized lengthy and realistic text reviews adapted from real-world examples, it is impossible to control for all extraneous attributes that might have differed across conditions. Therefore, Study 2 utilizes a different type of feedback format – seller ratings – that is not subject to this concern.

Lastly, intention was measured only at the end of the procedure, as it would be problematic to ask participants about their intent to purchase before all reviews had been presented. Therefore, H1b could not be tested. In Study 2, we remedy this issue by utilizing a design in which participants are presented information about pairs of treatment-control stores.

2.4 Study 2: Profile Ratings

Although not all real-world seller feedback forums provide text review information, most provide some form of seller ratings profile, depicting important criteria such as “on-time delivery”, “customer service”, etc. These aggregated profiles are often the first (or only) third-party information that will be utilized by potential customers. In addition to their widespread use, ratings are more easily processed than text reviews and more unequivocal. Importantly, the specific items that appear in ratings profiles often indicate matters of competence and/or integrity. At Bizrate.com, for instance, the profile items “variety of shipping options” and “order tracking” represent competence concerns, whereas the items “charges stated clearly before order submission” and “product met expectations” represent integrity concerns. From an experimental design perspective, these features allow for straightforward manipulation of information valence and domain while holding constant various potential confounds.

Given these methodological strengths, Study 2 utilized seller ratings as the experimental stimuli, and participants were presented with peer ratings of multiple potential sellers on a variety of items related to seller competence and integrity. In order to identify biased processing of the ratings, participants were exposed to pairs of ‘treatment’ and ‘control’ sellers. Treatment sellers had received both high and low ratings in the selected dimension, while control sellers had received neutral ratings on that dimension. Therefore, disproportionate weighting of positive and negative information would be indicated by evaluations that differed between treatment and control. In summary, the study utilized a 2 (item type: competence, integrity) \times 2 (seller type:

treatment, control) within-subject design to test whether negative ratings are more impactful than positive ratings, and (more importantly) whether this differential impact depends on the aspect of the seller being rated.

2.4.1 Method

2.4.1.1 Stimulus Materials

Six rating items appropriate for the study (three integrity, three competence) were identified by consulting actual websites and the prior trust literature. In order to be selected, it was important that: 1) all items correspond uniquely to either competence or integrity (with minimal overlap), and 2) the three items in a dimension cover distinct aspects of that dimension. During a series of pretests, undergraduate students were asked the extent to which various candidate items related to a seller's competence and integrity, using 7-point scales (1 = "not related," 7 = "highly related"). After pretesting and refinement, the set of items described in Table 3 was selected for inclusion.

Table 3: Items Used in Store Profiles

Section	Categories
Competence	On Time Delivery
	Order Processing and Tracking
	Website Design and Ease-of-Use
Integrity	Straightforward Return Policy
	Product Described Truthfully
	Charges Clearly Stated




In place of numeric scales, profile ratings were provided on 'Smiley' scales like those commonly used in online feedback forums (e.g., Bizrate.com), as illustrated in

Figure 2. Originally used for children and primary school students, smiley scales have adequate reliability and validity (Davies and Brember 1994; O'Rourke 2004). Moreover, this choice was based on two primary reasons: first, the smiley scale has a clearly identifiable neutral point, whereas it is unclear whether, e.g., 'three stars' is perceived as neutral on a five-star rating scale. Second, whereas numeric scales may prompt unmotivated participants to simply calculate an average rating instead of focusing on the individual items, the numeric meaning of smiley ratings is less obvious and less conducive to such simplifications.

2.4.1.2 Procedure


Thirty-five undergraduates at a large university participated in the study. This sample size was deemed adequate because this study is completely within-subjects, so each participant was exposed to all conditions and served his/her own control. The cover story (buying a digital camera) was similar to that of Study 1, except that participants were asked to assume that they had tentatively selected four online retailers and were now deliberating among them. Participants were then presented with four different 'profiles' that summarized recent customer ratings of the four sellers. The four seller profiles were each composed of two sections, one containing items related to competence and the other containing items related to integrity (see Figure 2 for an example). Each section listed three relevant items, along with a rating for each item. After viewing both sections, participants completed measures of trust and intention to purchase from the seller. At the end of the session, participants completed a manipulation check, described below.


Store A Profile - Competence Items


ITEM	AVERAGE RATING
On Time Delivery	
Order Processing and Tracking	
Website Design and Ease-of-Use	


Based on the information above, how **incompetent/competent** would you say that this store is overall?

Very Incompetent


☐


☐


☐


☐

Very Competent



☐

Figure 2: Screenshot of Profile’s Competence Items

The item type factor was manipulated through the items displayed in the first section of a profile. The items displayed in the first section reflected the targeted information domain (competence or integrity), and the type of items in the second section reflected the non-targeted information domain.

The seller type factor was manipulated through the ratings assigned to the targeted items. Profile ratings were represented by a 5-point smiley scale ranging from “very negative” to “very positive”; Table 4 depicts (numerically) the specific ratings assigned to each store. Treatment and control stores differed only in the set of ratings presented in the first (target) profile section: for control stores, these three items were all given neutral ratings, while for treatment stores, the first item was rated neutral and the last two items

were rated equally extreme but opposite in valence (very negative and very positive). The specific items rated positively and negatively were counterbalanced, ensuring that any observed effects could not be attributed to the importance or strength of individual items. Ratings in the second (non-target) profile section were identical for treatment and control stores. As a result of this design, differing perceptions of treatment and control stores would indicate differential weighting of negative vs. positive information.

Table 4: Ratings Used in Store Profiles

		Treatment Group	Control Group
Store Profile in Competence Condition	1 st Section (Competence)	(0, -2, 2)	(0, 0, 0)
	2 nd Section (Integrity)	(2, 1, -1)	(2, 1, -1)
Store Profile in Integrity Condition	1 st Section (Integrity)	(0, -2, 2)	(0, 0, 0)
	2 nd Section (Competence)	(2, -1, 1)	(2, -1, 1)

Note: The numbers in parentheses reflected the ratings' value, but the subjects saw smiley scales.

All participants viewed the same four profiles, and the order of presentation was counterbalanced. In order to disguise the purpose of the study, no treatment store ever appeared adjacent to its corresponding control store.

2.4.1.3 Measures

After viewing all the information about each store, participants were asked about their trust and intention to purchase from the seller. Measurement of trust was performed using the same scales used in the previous study. Intention to purchase was measured

with three items adapted from Jarvenpaa et al. (2000) and Pavlou (2003); these items are depicted in Appendix A.

2.4.2 Results

2.4.2.1 Manipulation Check

In the final portion of the study, participants were given formal definitions of competence and integrity, and then asked to rate the extent to which each of the six profile items was related to these dimensions on a 7-point scale (1 = “Not Related” and 7 = “Highly Related”). Confirming that the items successfully targeted their relevant dimensions, a paired-sample t-test indicated that competence items were perceived to be more related to competence than to integrity ($M = 5.61$ vs. 3.44 , $p < 0.001$), and integrity items were perceived to be more related to integrity than to competence ($M = 6.28$ vs. 5.17 , $p < 0.001$).

2.4.2.2 Trust and Intentions

H1 was tested by comparing trust and intention to purchase for treatment and control stores. Indices for trust and purchasing intention were formed by averaging participants’ answers to the corresponding questions (Appendix A), and the pattern of means for these variables is shown in Figure 3. A repeated-measures ANOVA was conducted on trust and intentions, with profile dimension and store type included as fixed factors. In both cases, analyses revealed a significant interaction, such that the difference in treatment vs. control stores depended upon the dimension involved ($p < 0.001$). In keeping with the negative bias observed for perceptions of seller integrity, follow-up

comparisons revealed that trust in the integrity treatment store was significantly lower than trust in the integrity control store ($M = 2.62$ vs. 3.85 , $p < 0.001$). Furthermore, intention to purchase from the integrity treatment store was significantly lower than intention to purchase from the integrity control store ($M = 2.62$ vs. 3.80 , $p = 0.001$). On the other hand, trust in the competence treatment store did not differ significantly from trust in the competence control store ($M = 3.67$ vs. 3.85 , $p > 0.3$), and neither did intention to purchase ($M = 3.49$ vs. 3.88 , $p > 0.1$). Taken together, these results support H1a-H1b and provide further evidence that the overweighting of negative information is domain-specific; negativity bias was more pronounced for integrity-related information than competence-related information.

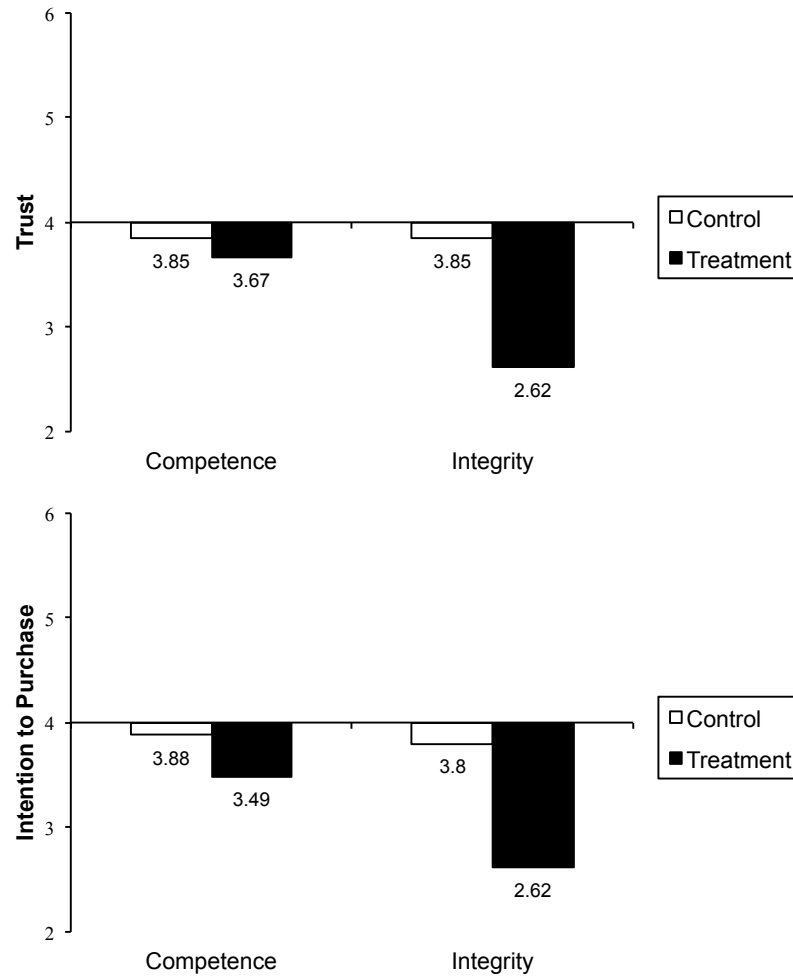


Figure 3: Asymmetric Negativity Biases in Trust and Purchasing Intention

2.4.3 Discussion

By extending our investigation to seller profile ratings, Study 2 provided a replication and extension of Study 1. The use of non-verbal ratings enabled straightforward manipulation of information valence and domain without introducing extraneous factors. Results converged strongly with those of the previous study and with our theoretical arguments: negativity bias was not universal, but rather depended on the

type of seller information being evaluated. Specifically, the bias regarding trust and actual intentions to transact with the sellers was more pronounced for information related to integrity than for information related to competence.

Studies 1 and 2 explored negativity bias by presenting a piece of positive information and a piece of negative information simultaneously, and presented direct evidence for negativity dominance. These studies provided converging evidence for our first hypothesis, but did not allow investigation of H2. Therefore, in Study 3 we extended our exploration to the dynamics of trust recovery by examining the common situation in which one piece of negative evidence (review) is paired with multiple pieces of positive evidence. To enhance realism and decrease the complexity of the procedure, the third study added additional, ‘filler’ stores, utilized a star-rating format commonly used by review websites, and presented all profile information on a single screen.

2.5 Study 3: Trust Recovery

Study 3 was designed to examine the dynamic process of trust recovery as predicted in H3 in a common scenario, where one negative review is paired with multiple positive reviews. This study utilized a 2 (profile dimension: competence, integrity) \times 5 (positive ratings: 0, 1, 2, 3 and 4) mixed design, with profile dimension a between-subjects factor and number of positive ratings a within-subject factor. Similar to Study 2, participants were asked to assume that they had tentatively selected several online retailers and would now be reading the rating profiles of each retailer one by one before making a decision. Each store’s profile contained five items related to either competence or integrity. Participants reported their trust and purchase intention after reading each

store's profile. Therefore, by comparing across stores, one may examine the quantity of positive ratings needed for trust 'recovery', and whether this quantity differs for competence vs. integrity violations.

2.5.1 Method

2.5.1.1 Stimulus materials

We first utilized rounds of pretesting to identify two sets of five items (one set competence, the other integrity) that would be used in the store profiles. Similar to in Study 2, undergraduate students rated the extent to which various candidate items related to a seller's competence and integrity. After pretesting and refinement, the final items are listed in Table 5.

Table 5: Items Used in Store Profiles

Section	Categories
Competence	Website Design and Ease-of-Use
	Order Processing
	Variety of Shipping Options
	Product Packaging
	Order Tracking
Integrity	Product Described Truthfully
	Charges Clearly Stated
	Product Shipped as Promised
	Correct Product as Described
	Straightforward Return Policy

2.5.1.2 Procedure

Forty-eight undergraduates participated in this study. The cover story was similar as that of Study 2, except that the number of stores under consideration was ten. Before

the task, participants were randomly assigned to the competence or integrity group and presented with the definition of competence or integrity. Participants were also shown the profile of a practice store to help them understand and prepare for the task.

Participants then read the profiles of each store, one by one, on separate screens. Five of the stores represented our treatment manipulation, and five others were used as fillers. Unlike Study 2, profiles in this study consisted of a single section with five items: depending on their condition, participants saw only items pertaining to competence or integrity. We counterbalanced the sequence in which the five items were displayed. Beside each of the five items was either a corresponding rating for that item or the expression “N/A” (as shown in Figure 4). All ratings were presented using a five-point scale ranging from one star to five stars. To create a highly negative initial impression, the first item in the treatment stores was always rated one star. We manipulated the number of subsequent, positive items by assigning five star ratings to 0-4 of the remaining items in the profile; all other items received “N/A”. The order of presentation was constructed such that filler stores were presented between each treatment store.

Store B's Profile	
Product Described Truthfully:	★ ☆ ☆ ☆ ☆
Charges Clearly Stated:	★ ★ ★ ★ ★
Product Shipped as Promised:	★ ★ ★ ★ ★
Correct Product as Described:	N/A
Straightforward Return Policy:	N/A

Figure 4: Screenshot of Profile’s Items for Participants in Integrity Condition

2.5.1.3 Measures

After participants read each store's profile, they rated trust and intention to buy from that store. All constructs were measured as in Study 2, except that we only selected 2 items for trust and 2 items for purchasing intention to limit the number of total questions.

2.5.2 Results

2.5.2.1 Manipulation Check

In a separate pretest, 43 participants were asked to rate the extent to which items in the competence and integrity profiles were related to each dimension. A one-sample T-test indicated that the competence profile was considered to be more related to competence than to integrity ($M = 6.81$ vs. 2.53 , $p < 0.001$), and integrity profile was considered to be more related to integrity than to competence ($M = 6.88$ vs. 4.23 , $p < .001$). Thus, the manipulation of profile dimension was deemed to be successful.

2.5.2.2 Trust Recovery Results

Our main hypothesis concerned the ease with which impaired trust and purchasing intention could be recovered due to additional, positive information. In the absence of additional positive information (0 positive items), trust in the treatment store and purchase intention were very low (below 2.5). We conducted a 2 (profile dimension: competence, integrity) \times 5 (positive ratings: 0, 1, 2, 3 and 4) ANOVA on trust and

purchasing intention for the ten target stores. The number of positive ratings was entered as a within-subject factor. This analysis revealed a significant main effect of positive ratings ($p < 0.001$). As shown in Figure 5, trust and purchasing intention in both conditions progressively shifted upward from the minimum point established in the store with 0 positive reviews. Therefore, after participants saw a bad rating, it was possible for their trust to recover back to above neutral.

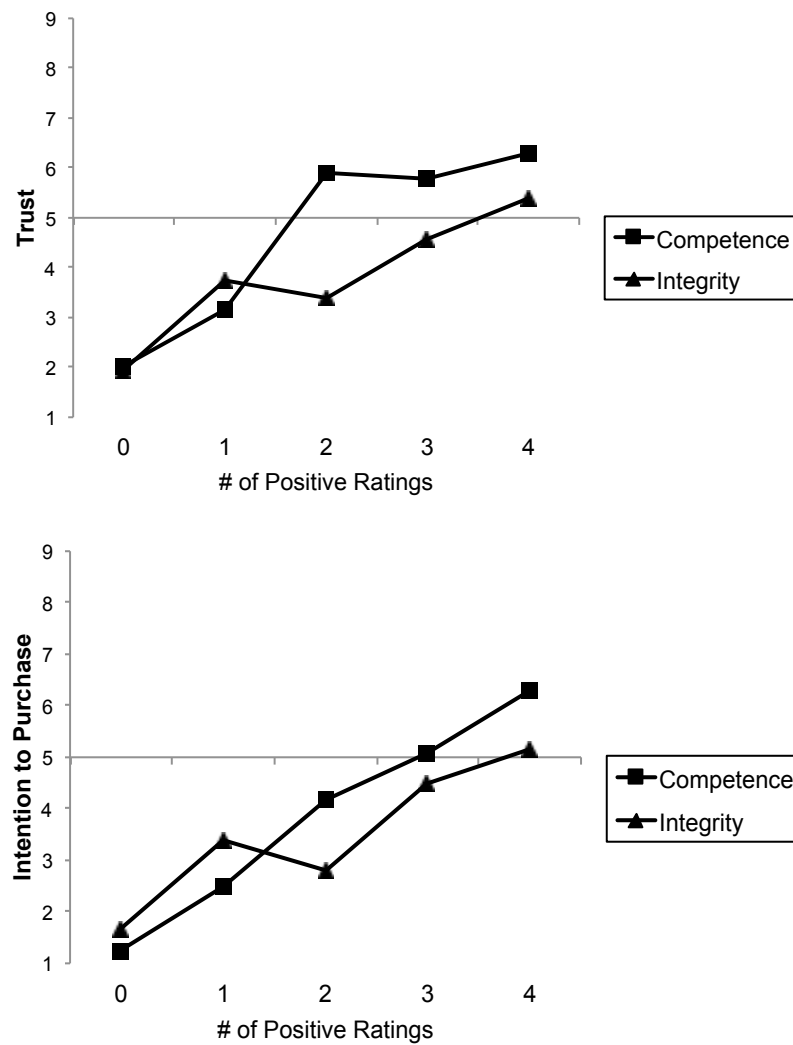


Figure 5: Trust and Purchasing Intention by Profile Dimension

More importantly, the ANOVA also revealed a significant interaction term between profile dimension and the number of positive ratings ($p < 0.05$), indicating that the pattern of recovery was different for participants in the competence and integrity conditions. Figure 5 revealed that for trust, approximately two positive items were required to cross the neutral point in the competence condition, while approximately four positive ratings were required in the integrity condition. For purchasing intention, the required number of positive items was approximately three in the competence condition and four in the integrity condition. Thus, the extent to which impressions formed from negative ratings were resistant to change was not the same for the participants in two conditions. Negative ratings regarding integrity were more difficult to contradict than those regarding competence. These results are consistent with Hypothesis 2.

2.5.3 Discussion

Study 3 extended the prior studies by exploring the dynamic trust recovery process in a setting where a negative rating appears with subsequent positive ratings. We proposed in Hypothesis 2 that although the impaired trust and purchasing intention arising from negative review information can be recovered, integrity perceptions are more resistant to change than competence perceptions. In order to test this hypothesis, we employed profile ratings of different stores as stimuli and varied the number of positive items that appeared after an initial negative item. Results indicated that impaired trust and purchasing intention can indeed be recovered with a sufficient number of positive ratings,

but the recovery process is faster for competence than for integrity. In sum, H2a and H2b were supported.

Note that both trust and purchasing intention recovered to above neutral after 4 positive ratings, and it is common in the real settings that the number of positive reviews far outweigh that of negative reviews for a retailer. However, this finding by no means indicates that negativity bias is not a concern for sellers: recovering trust and purchasing intention to above neutral does not mean that consumers will commit themselves to an actual purchase. Instead, they may wait and search for other better retailers. In other words, the findings reported in this study imply that ease of trust recovery after a violation depends on the domain of review information, but sellers should not treat the recovery process to be an easy one.

2.6 General Discussion

Online feedback forums represent a powerful means by which consumers may obtain valuable information about potential sellers. Given the vast amount of information available and the elevated risk inherent to e-commerce, it is not surprising that consumers may exhibit a negativity bias in processing this information; indeed, the idea aligns well with accepted principles of cognitive processing, and consistent results have been reported in prior literature. However, the findings presented in this paper suggest that a more nuanced framework is necessary for understanding the interpretation of seller feedback and the recovery process after a trust violation.

Across three experimental studies utilizing distinct stimuli and methodologies, findings supported our argument that the domain of information conveyed in a seller

review influences the extent to which processing of that review is biased. Studies 1 and 2 utilized text reviews and ratings as stimuli, respectively, and created the potential for capturing negativity dominance by simultaneously presenting positive information and negative information. Both studies provided evidence of a stronger negativity bias for information concerning seller integrity than that concerning seller competence, a result that was obtained across both trust and purchase intention. Study 3 extended our investigation into the dynamics of trust recovery: specifically, we examined the extent to which initially negative impressions caused by poor feedback ratings are resistant to change, and whether this resistance is moderated by the seller traits implicated in the feedback. Results revealed that although impaired trust and purchasing intention were recoverable, negative impressions caused by dishonesty were more resistant to change than those caused by incompetence. Together, the converging studies support our theoretical argument based on the diagnostic value of positive vs. negative information. Moreover, the findings make us confident in concluding that negative reviews are not all “created equal.” Rather, the axiom that “bad outweighs good” is most applicable to reviews containing information about a seller’s moral standing; when review information refers to seller competence, negativity bias will be attenuated, or even eliminated.

2.6.1 Theoretical Implications

The impact of negativity bias on sales and price premiums has been established previously (Ba and Pavlou 2002; Chevalier and Mayzlin 2006). Although prior studies have provided valuable insights using cross-sectional field study designs, these approaches are hindered by lack of control and the constraints of secondary data, and

they are unable to provide evidence for negativity dominance. We conducted two controlled experiments (Study 1 and 2) in which the valence of information was manipulated by text reviews or ratings, while holding information extremity constant. Utilizing carefully pretested stimulus materials, we were able to effectively capture negativity dominance as it arose and, more importantly, compare its impact on trust and intentions across information domains. This experimental approach represents a valuable supplement to other methods of inquiry used in the study of buyer-seller relationships online.

Our research explored not only on the presence of negativity bias in online word-of-mouth, but also its magnitude under various conditions. In contrast to recent work examining bias in terms of word-of-mouth volume (Khare et al. 2011) or observational learning (Chen et al. 2011), we focused directly on the content being shared. In particular, we applied insights based on information diagnosticity and attribution theory to explain the distinct natures of competence and integrity, which are fundamental dimensions of social perception. In e-commerce settings, positive reviews far outweigh negative reviews in quantity, yet the greater weight of negative reviews (i.e., negativity bias) may be contingent on how observers interpret and make sense of the reviews. Our results suggest that although negativity bias does influence trust in this setting, it is much more pronounced for information regarding seller integrity than that regarding seller competence. These results contribute towards an established need for broader understanding of consumers' trust formation and decision-making processes (Ahluwalia 2002). Looking ahead, the negativity bias observed in prior empirical studies might be reconsidered in light of the qualifications presented here. An examination of diagnosticity

could be instrumental in explaining how various kinds of information regarding products, sellers, or purchase context are weighted and integrated in consumers' judgment and decision-making.

Most prior studies have treated trust as a static state, whereas the formation and recovery of trust is unquestionably dynamic in nature (Mayer et al. 1995). Thus, it is worthwhile and meaningful to evaluate trust from a dynamic perspective, and Study 3 presented one approach by exploring the recovery process after a trust violation. Our findings imply that effects of a negative review are indeed counteracted by a sufficient number of positive reviews, but, more importantly, a negative integrity-related review is more detrimental than a negative competence-related review and requires more counterevidence for recovery. Important in itself, this example also illustrates the need for greater understanding of how trust is updated with accumulating knowledge.

2.6.2 Managerial Implications

Seller feedback forums (BizRate.com, Angieslist.com, etc.) are increasingly consulted by consumers as part of their typical shopping patterns. Naturally, because negative ratings or reviews often carry more weight than their positive counterparts, it is especially critical to avoid them in the first place; that is, time and expense spent preventing customer dissatisfaction pays additional dividends by enhancing future shoppers' perceptions of the firm. This implication is consistent with the conventional wisdom that on average, a satisfied customer tells 3 people about a good product experience, whereas a dissatisfied customer gripes to 11 people (Kotler and Armstrong 2010). On the other hand, our research demonstrates that customers may be especially

sensitive to negative (vs. positive) word-of-mouth dealing with a seller's integrity rather than a seller's competence. That is, a review describing delivery delays, order-tracking errors, etc. may indeed undermine trust formation, but a review describing deceptive sales tactics, overcharges, etc. provides a powerful foundation for distrust. This premise carries a number of important implications: for example, in certain cases, either competence or integrity is a viable explanation for a transgression (e.g., a product description is inaccurate). In these cases, vendors may be well served to explain the issue as a severe 'blunder' rather than a moral transgression. Furthermore, to the extent that customer service resources are limited, firms may choose to concentrate those resources on handling integrity concerns proactively (e.g., resolving complaints of dishonesty), before incriminating reviews are posted.

Negative reviews are often unavoidable, and they may be spread across various third-party review sites. Study 3 demonstrated that although the influence of negative ratings on trust can be counteracted, a substantial amount of counterevidence is required. In settings where reviews do not accumulate fast enough, sellers might consider providing a public response explaining the incident or contacting dissatisfied customers directly in an attempt to earn forgiveness and (ideally) retraction of the critical review (Kim et al. 2004). This option deserves particular consideration in the case of integrity-related criticisms, which can be especially difficult to overcome.

2.6.3 Limitations

Our studies provided essential evidence for negativity dominance that was mostly neglected in prior literature, while the number of reviews in real online settings is

generally large. Even though we investigated cases in which four or five ratings were encountered in Study 2 and 3, real retailers can easily accumulate hundreds of, and even thousands of, ratings and reviews. Evidence for negative potency for such scenarios was abundant, but whether negative potency as exemplified in such large scales also depends on the domain of seller behaviors implicated in the reviews merits future research.

Second, our hypotheses were grounded in impression formation literature and attribution theory, and the three experiments provided direct evidence for the hypotheses. However, we did not measure the potential mediators and were thus unable to test the mediating processes. Although we believe that this paper contributes to the negativity bias literature by examining negativity dominance, future research can extend our studies and explore the potential mediators that can explain the effect of social domain on trust and purchase intentions.

Third, our major focus is to demonstrate negativity dominance in online word-of-mouth, and we simply counterbalanced the sequence in which positive and negative information appears in the first two studies. However, the temporal dimension of the reviews is also an important factor in trust formation and recovery. Additionally, the negative rating always appeared first in Study 3, where we assume that people read reviews from the top, and they will develop a negative initial impression based on the top one. However, this may not always be the case, and having the negative review in other positions (e.g., in the end) may create different patterns of the results beyond our scope. Future studies can examine the temporal factors and explore this possibility.

2.6.4 Future Research

The possibility that negativity bias in trust perception is domain-specific also presents a number of unexplored questions. The current experiments were an especially conservative test of our hypotheses, given that participants incurred no actual risk of monetary loss, and we believe that real consequences would yield an even greater discrepancy in negativity bias across information domains. Future research might investigate this directly. Furthermore, decisions in the present studies were immediate, and participants viewed a modest amount of information about potential sellers. However, for purchases high in complexity, risk, or importance, consumers may seek vast amounts of information before making a decision. In these cases, memory for the relative strengths and weaknesses of different vendors plays a critical role, and a number of important questions present themselves. For example, what forms of feedback are most likely to be remembered (negative or positive, integrity- or competence-related, ratings or reviews, etc.), and for how long? How might firms help customers to encounter, remember, and retrieve favorable (but not unfavorable) feedback?

To extend the current research, a number of other relevant variables present themselves. Although we examined the dynamic process of trust recovery (in Study 3), the dynamics of trust formation itself also merit exploration. Our studies demonstrated that the magnitude of negativity bias depends on information domain, but it would be interesting to examine whether this effect holds over time. Furthermore, in addition to peer reviews and ratings, numerous other information sources convey the competence and integrity of online sellers: e.g., website quality and complexity may signal competence, while the presence or absence of third-party seals may signal integrity. We

suspect that these information sources are susceptible to the same pattern of biases observed in the present research, but this issue merits further exploration.

Rather than waiting for a single negative review to be offset by countervailing positive reviews, vendors often take a more active role. For example, a seller might contact the unsatisfied customer, clarify misunderstandings, solve the underlying problems, and/or provide remedial compensation (Kim et al. 2006). Which combination of remedial strategies is most effective in recovering trust? Are the effects of these strategies contingent on characteristics of the seller and buyer? Exploration of these questions would broaden the present investigation while informing online communication strategies more broadly.

2.7 Conclusion

At the presence of abundant empirical evidence suggesting negativity bias in online word-of-mouth, our paper contributes to this literature by examining negativity dominance, the most robust demonstration of negativity bias, and by exploring its boundary conditions. Our research provides experimental evidence that negative seller reviews are not ‘created equal,’ and that the universality of negativity bias in this setting has been exaggerated. As such, we believe that this work extends current understanding of negativity bias, and we look forward to further research exploring negativity dominance and its boundary conditions in online word-of-mouth.

CHAPTER 3

SHOULD YOU BE FIRED UP IF YOU WANT TO BE HELPFUL?

EFFECTS OF EMOTIONAL AROUSAL IN ONLINE WORD-OF-MOUTH

3.1 Introduction

Modern consumers are increasingly consulting peer opinions rather than firms themselves for information pertaining to their purchase decisions (Bickart and Schindler 2001). Due to online communication platforms, consumers across the globe can write online reviews for products and services, which exert considerable influence on future consumers (Chevalier and Mayzlin 2006; Forman et al. 2008). Various third-party platforms have accumulated large amounts of reviews for prospective customers: product reviews at Amazon, retailer reviews at eBay, movie reviews at IMDb, software reviews at CNET, etc.

In this paper, we examine antecedents of the perceived helpfulness of online reviews in the app market. In recent years, a new type of reviews has become available for apps on mobile devices, as pioneered by Apple. Debuted on July 10, 2008, the App Store is a digital application distribution platform for the iPhone and other iOS devices (Apple 2008). Apple enables existing users of an app to leave a rating and detailed feedback for the app. Given the difficulty in deciding among a large quantity of apps for any category in this highly competitive market, the review system adopted by the App Store is a powerful channel for informing prospective users and driving sales. Not surprisingly, some ill-intentioned app developers have tried to game the App Store's

ranking system by manipulating app reviews (Kaneshige 2010), and Apple has recently warned developers against such endeavors (Apple 2012).

We define review helpfulness in this context as the extent to which a peer-generated evaluation of an app is perceived by prospective users to facilitate their decision-making process (see Mudambi and Schuff 2010). Like many other third-party platforms that host online reviews, Apple provides the opportunity for review readers to evaluate helpfulness directly by answering the question “Was this review helpful?” The number of users who clicked “Yes” out of the total number of users who casted a vote is also displayed below each review. In addition, users can sort reviews by displaying the “Most Helpful” reviews on top.

Understanding the determinants of review helpfulness perceptions is an important concern for users, developers, and review platforms. The ability to evaluate review helpfulness helps prospective consumers to cope with information overload and assists their decision process before making a purchase (Cao et al. 2011; Mudambi and Schuff 2010). Promoting the most helpful reviews not only informs developers of prior customers’ most urgent concerns (Chen et al. 2008), but also benefits review platforms themselves by attracting attention and increasing customer ‘stickiness’ (Connors et al. 2011; Dabholkar 2006). However, considerable time is required to obtain useful feedback from peer-voting mechanisms and other direct measures (Zhang and Tran 2010) and understanding the determinants of review helpfulness could speed up the process. Such knowledge would not only guide developers in dealing with user reviews, but could also help review platforms design writing guidelines to enhance the overall quality of reviews.

Prior research exploring antecedents of review helpfulness has focused mostly on cognitive factors, such as numerical ratings, reviewer location, and product type (Chevalier and Mayzlin 2006; Forman et al. 2008; Mudambi and Schuff 2010). A few scholars have instead examined the actual textual content of reviews, revealing that their content has considerable influence on helpfulness, over and above the effect of ratings (Cao et al. 2011). However, these studies have restricted their analysis to semantic characteristics, neglecting the potential effect of affective processes.

In reality, reviewers commonly convey their emotions in text reviews (Hennig-Thurau et al. 2004), and this emotional content is useful for readers to judge review quality and make better decisions (Cao et al. 2011; Kuan et al. 2011). In particular, emotion is often characterized within a two-dimensional space of valence and arousal (Russell 1980). Regarding valence, prior empirical studies have repeatedly observed a negativity bias regarding numerical ratings, and it is very likely that the bias will also apply to the emotional content. However, the effect of emotional arousal on review helpfulness has not been explored. To address this gap, we ask the following research questions: How does the emotional arousal of reviewers influence the perceived helpfulness of reviews? What are probable mechanisms underlying this effect? Drawing from dimensional theories of emotion and the Yerkes-Dodson law of cognitive performance, we propose that the amount of arousal conveyed in a text review will influence evaluations of that review in a non-linear fashion, such that beyond a certain optimal level, higher arousal reduces the perceived helpfulness of a review. Furthermore, we suggest that the detrimental effect of arousal is present for negative reviews even when objective review content is controlled for, and that heightened perceptions of

irrationality underlie this effect. We conducted two studies using convergent methods to test these hypotheses.

3.2 Literature Review and Hypotheses

3.2.1 Affect and Emotion

In psychological research, the term ‘affect’ refers to a general category of mental processing that involves subjective internal feelings (Cohen et al. 2008). In contrast, ‘cognition’ often refers to the mental processes involved in thinking, remembering, judging and problem-solving. There was a long debate among psychologists concerning whether affect and cognitions are independent, and whether affect can occur without cognition (Lazarus 1982; Zajonc 1980; Zajonc 1984). We adopt the generally accepted position that cognition and affect are interdependent. In particular, we believe that affective processing is often inherently cognitive (Solomon 2008), and affective states can influence decision-making through cognitive processes (Forgas 1995).

‘Emotion’ constitutes a specific category of affective processes, defined here as “a mental state of readiness that arises from cognitive appraisals of events or thoughts” (Bagozzi et al. 1999, p. 184). Emotion differs from ‘mood’ in that the former tends to be more intense and brief, and more context specific with a particular cause (Ekman 1992; Frijda 1993). Therefore, we focus on emotion in this paper, because the reviews in the app market are targeted at specific apps.

Numerous researchers have suggested fundamental dimensions for the classification of emotions (Brosch et al. 2010; Mano 1991; Watson and Tellegen 1985). Among these dimensions, valence and arousal have been consistently identified as most

important (Niedenthal 2008; Russell 1980). Valence describes the extent to which an experience is pleasant or unpleasant, while arousal describes the extent to which an actor is activated or deactivated (Niedenthal 2008). Arousal ranges from ‘calming/soothing’ to ‘exciting/agitating’ (Heilman 1997). Emotions of the same valence may induce different levels of arousal (Berger and Milkman 2012): for example, anxiety, anger, and sadness are all negative emotions. However, anxiety and anger are characterized by heightened arousal, while sadness is characterized by low arousal (Barrett and Russell 1998).

3.2.2 Emotional Arousal and the Yerkes-Dodson Law

As a starting point, we assume that a user has used an app for an extended period and plans to write a text review. A fundamental contention of this paper is that the emotional arousal of reviewers is likely to interfere with their thinking and performance during the review process. According to the Yerkes-Dodson law (1908) described below, we propose a curvilinear relationship between the level of emotional arousal and perceived helpfulness of reviews.

At low levels of emotional arousal, there are multiple reasons to expect that an increase in arousal will result in better performance. Individuals at very low levels of arousal are characterized by emotions such as relaxed, calm, or bored (Yik et al. 2011). Task performance suffers in this context because individuals devote little energy to information processing, exert minimal effort on the task, and focus attention on other matters (Kahneman 1973). As one’s arousal level increases, added emotional intensity can evoke increased effort in information processing, resulting in higher performance. In

summary, when the arousal level is low, an increase in arousal can improve task performance.

Beyond moderate levels of arousal, however, additional arousal is likely to have a detrimental effect. It is generally assumed that as “cognitive misers,” individuals rely on a central pool of resources for information processing and cognitive tasks (Krueger and Funder 2004; Tversky and Kahneman 1974). However, these resources are finite, and when they are exhausted, performance will suffer (Johnston and Heinz 1978; Wickens 1991). In particular, there is ample evidence that elevated arousal impairs cognitive capacity, narrows attentional focus, and disrupts information processing (Fedorikhin and Patrick 2010; Sanbonmatsu and Kardes 1988; Shapiro et al. 2002). Under high levels of arousal, therefore, its cognitive costs result in diminished performance.

We suggest that the relationship discussed above is especially likely to occur in the app market context. Reviews represent a form of discursive writing, which is deliberative in nature (Vygotsky 1964). For example, users often need to reconstruct their experience with the app, evaluate pros and/or cons, and translate them to the verbal form that readers will understand. Therefore, it is reasonable to expect that some effort, and thus arousal, is need for constructing a helpful review. However, beyond a certain point, additional arousal impairs a reviewer’s cognitive processing capabilities and leads to less helpful reviews. In line with abundant evidence confirming the Yerkes-Dodson law in literature (Buck 1988; Yates 1990), therefore, we expect the effect of emotional arousal to follow an inverse U-shaped curve: a moderate level of arousal results in the most helpful reviews, while very low and very high arousal levels lead to less helpful ones. We propose the following hypothesis:

Hypothesis 1: The emotional arousal in a text review is associated with its perceived helpfulness according to an inverse U-shape: (a) for reviews containing low levels of arousal, additional arousal is perceived to be beneficial, and (b) for reviews containing high levels of arousal, additional arousal is perceived to be detrimental.

In Study 1, we tested H1 in a realistic context by examining actual reviews collected from Apple's App Store. We collected over 1 million reviews of approximately 40 thousand apps, coded those reviews for arousal using text analysis tools, and examined the relationship between arousal and review helpfulness.

3.3 Study 1: App Store Reviews

Our primary goal in this study was to utilize secondary data to test H1 in a realistic, online, word-of-mouth setting. In particular, we collected actual review data from Apple's App Store, which launched in mid-2008 and now has over 500,000 apps approved by Apple. The App Store provides user ratings and reviews for apps, and the site had accumulated nearly two years of user reviews at the time of data collection. When leaving a review, users can evaluate the app by leaving a 1-to-5 star rating and writing a text review to describe their experience with it (see Figure 6). The review page of an app lists all reviews in reverse chronological order.

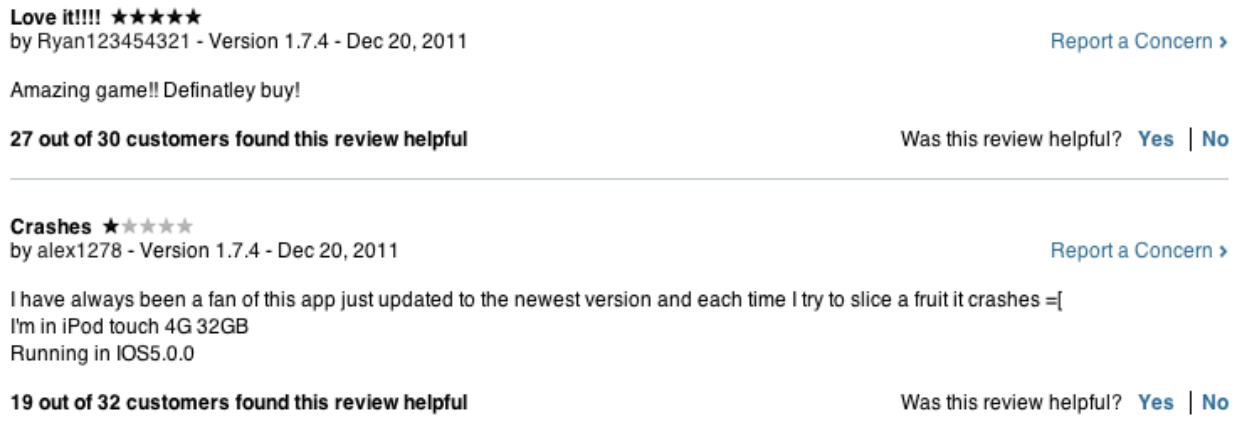


Figure 6: Screenshot of Two App Reviews for “Fruit Ninja”

3.3.1 Data Collection

We collected the data in April 2010, using individual reviews as the unit of analysis. We began by identifying 62,266 apps that appeared in the top 500 rankings of all app store categories for the first three months of 2010 (20 in total, including games, business, reference, social networking, etc.). Among the identified apps, 40,417 had at least one review, and we retrieved all these historical reviews. For each review, we collected the following information: rating, text review content, helpful votes, and total votes. We also collected app-level information, including the average rating, count of all ratings, category, and whether or not the app is paid.

Next, we performed the following steps for data cleaning before further analyses. First, we dropped 94,815 reviews that included non-English characters. Next, we dropped 2,743 reviews that did not include any text content and 38 reviews that were rated 0 (probably due to system errors). After these steps, we obtained a set of 1,623,497 reviews. Of this set, 418,415 reviews (over 25%) had received at least one helpfulness vote. Analysis was conducted on these reviews.

3.3.2 Variables

Review helpfulness, the dependent variable of interest, was operationalized as follows. Below each review, iTunes lists the question “Was this review helpful?” along with “Yes” and “No” options. A review that has received at least one vote will display the number of “helpful” votes and total votes immediately below the review content. Helpfulness was measured as the number of people who voted “Yes” divided by the total number of people who cast a vote, ranging from 0 to 1. A higher percentage indicates a more helpful review. The average helpfulness of the analyzed reviews was 0.59, indicating that most reviews in the final set were considered relatively helpful. Tables 6 and 7 present a summary of statistics and correlations for this and the remaining variables (described below).

Table 6: Descriptive Statistics for Final Review Pool*

	Variable	Mean	Std. Dev.	Min	Max
1	Review helpfulness	0.59	0.42	0	1
2	Rating	3.45	1.68	1	5
3	Length	41.63	48.96	1	1134
4	Reading difficulty	8.70	54.60	-16.1	25428.7
5	Emotional valence	1.89	0.14	1	3
6	Emotional arousal	1.68	0.12	1	3

Notes:

* Refer to Table 8 for the definitions of these variables.

Table 7: Variable Correlations for Final Review Pool (N = 414,336)*

Variable	1	2	3	4	5	6
1 Review helpfulness	1					
2 Rating	0.363	1				
3 Length	0.132	0.027	1			
4 Reading difficulty	0.047	0.037	0.013	1		
5 Emotional valence	0.101	0.327	-0.129	0.013	1	
6 Emotional arousal	0.002	0.103	-0.162	0.088	0.376	1

Notes:

* Among the set of the final reviews, 414,336 reviews (99%) got values for emotional valence and arousal.

Emotional valence and arousal of the text reviews were measured using Whissell's (2009) text analysis software, Revised Dictionary of Affect in Language. It has been widely used for quantifying emotion dimensions in psychology and linguistics, and there was evidence supporting its reliability and validity (Whissell 2009). Originally designed to quantify the valence and arousal of distinctly emotional words, this software was revised to enhance its applicability to analyzing text samples of natural language. The revised dictionary contains 8,742 words characteristic of natural language, and it has been shown able to match 9 out of every 10 words in most language samples. Importantly, the software provides scores for emotional arousal. When a text sample is scored, the software will match its words with its dictionary; whenever a match is found, the dictionary values for valence and arousal will be retrieved. The mean of all values for a particular dimension is calculated, representing the final score for that dimension in the text sample. We employed this text analysis tool to measure the emotional valence and arousal of the app reviews in our data set. Among voted reviews in the set, 99% received values for valence and arousal.

In the following analyses, we controlled for variables that have been used in prior literature examining antecedents of review helpfulness (Korfiatis et al. 2008; Mudambi and Schuff 2010). These variables include review rating, length, reading difficulty, and certain app characteristics. (1) The review rating refers to the star rating of a review; the more stars a review received, the more positive the review is. Rating ranged from 1 star to 5 stars, and the average rating for the reviews in the set was 3.45. (2) Review length was operationalized as the number of words in a review; a longer review often provides more information and thus is deemed more helpful. The analyzed reviews had on average 41.63 words. (3) We calculated the Coleman–Liau Index (CLI) to control for review reading difficulty. CLI is an estimate of the U.S. grade level that a student would need to have achieved in order to read and understand the text (Coleman and Liao 1975). On average, the reviews were written at a 9th grade level. (4) We also controlled for the effects of app characteristics, including an app’s average rating, the count of all its prior ratings, price, and category. Average rating captures the overall quality of an app, while the count of all ratings for an app captures its popularity. We used a price dummy, which is equal to 1 if an app was paid. Additionally, each app belonged to one of the twenty categories, so we added nineteen dummies to control for cross-category heterogeneity. The operationalization of all variables is summarized in Table 8.

Table 8: Variables and Operationalizations

Variable Type	Variable Level	#	Variable	Operationalization	Notes
DV	Individual Review	1	Review Helpfulness	# helpful_votes / # total_votes	Range: [0, 1]
IV	Individual Review	2	Emotional Valence	average valence score of identified words	Range: [1, 3] Coded by Revised Dictionary of Affect in Language
		3	Emotional Arousal	average arousal score of identified words	
Control	Individual Review	4	Rating	# of stars	Range: [1, 5]
		5	Length	# of words	
		6	Reading Difficulty	Coleman-Liau Index	U.S. grade level necessary to comprehend the text Range: [1, 5]
		7	Quality	average rating	
	Store	8	Popularity	# of ratings in total	Whether an app is paid The category (20 in total) an app belongs to
		9	Price	=1 if the app is paid	
		10	Categories	=1 if the app belongs to that category	

3.3.3 Data Analysis and Results

We used Tobit regression to analyze the final set of reviews (Mudambi and Schuff 2010) for the reasons below. First, the dependent variable was a ratio. Thus, it was bounded in range and censored in nature. Second, not every review reader casts a helpfulness vote, so a selection bias may exist. Therefore, ordinary regressions of this sample containing only voted reviews might produce biased estimates (Greene and Zhang 2003).

The results of our empirical analysis are shown in Table 9. All continuous independent variables were standardized to unify presentation of the results. As indicated in Model 1, coefficients of the control variables were mostly consistent with prior

literature. Review rating was significantly negatively related to review helpfulness ($\beta = -0.124, p < 0.01$), indicating a negativity bias. Moreover, a review was considered more helpful to the extent that it was longer ($\beta = 0.242, p < 0.01$), and more difficult to understand ($\beta = 0.005, p < 0.05$). A probable reason for the latter result is that more complex reviews give more information, resulting in heightened helpfulness, but they are also harder to read. Coefficients for average rating ($\beta = -0.068, p < 0.01$) and count of ratings ($\beta = -0.607, p < 0.01$) were negative. Therefore, controlling for other variables, reviews of high quality or popular apps were considered less valuable. Lastly, reviews of paid apps were more helpful than those of free apps ($\beta = 0.534, p < 0.01$) (presumably because paid apps carry higher risk).

Table 9: Tobit Analysis Results for Final Review Pool

Dependent variable: Review Helpfulness

	Model 1	Model 2	Model 3	Model 4	Model 5
Rating	-0.124*** (0.002)	-0.108*** (0.002)	-0.118*** (0.002)	-0.108*** (0.002)	-0.109*** (0.002)
Length	0.242*** (0.002)	0.232*** (0.002)	0.233*** (0.002)	0.230*** (0.002)	0.228*** (0.002)
Reading difficulty	0.005** (0.002)	0.119*** (0.010)	0.124*** (0.010)	0.122*** (0.010)	0.127*** (0.010)
Average rating	-0.068*** (0.002)	-0.059*** (0.002)	-0.062*** (0.002)	-0.059*** (0.002)	-0.060*** (0.002)
Number of ratings	-0.607*** (0.003)	-0.607*** (0.003)	-0.606*** (0.003)	-0.606*** (0.003)	-0.606*** (0.003)
Paid	0.534*** (0.004)	0.523*** (0.004)	0.524*** (0.004)	0.523*** (0.004)	0.521*** (0.004)
Category dummies	included	included	included	included	included
Emotional valence		-0.051*** (0.002)		-0.043*** (0.003)	-0.040*** (0.003)
Emotional arousal			-0.036*** (0.002)	-0.020*** (0.002)	-0.008*** (0.003)
Arousal ²					-0.013*** (0.001)
Constant	-2.144*** (0.021)	-2.134*** (0.021)	-2.136*** (0.021)	-2.135*** (0.021)	-2.122*** (0.021)
N	1623497	1587363	1587363	1587363	1587363
Log Likelihood	-1375268.8	-1358403.8	-1358510.8	-1358368.3	-1358281.1
Pseudo R-square	0.0672***	0.0658***	0.0657***	0.0658***	0.0658***

Notes:

Standard errors in parentheses

*** significant at 0.01; ** significant at 0.05; * significant at 0.1

Next, we examined Hypothesis 1 by entering emotional valence, arousal, and the squared term of arousal step by step in the Tobit regressions (see Table 9, Model 2-5²). The coefficient of valence was negative and significant, indicating that negativity bias also applied to the emotional content of app reviews. The coefficient of arousal was also negative and significant, indicating that more aroused reviews were considered less helpful on average. When the squared term of arousal was entered in Model 5, its coefficient was negative ($\beta = -0.013, p < 0.01$), indicating a non-linear relationship. The coefficient of arousal was still negative ($\beta = -0.008, p < 0.01$). Based on these coefficients, the relationship between arousal and review helpfulness is approximately depicted in Figure 7. As predicted by H1, when the arousal level is very low, more aroused emotions will increase the perceived helpfulness of a review. However, after an optimal level of arousal is achieved, more arousal will lead to lowered review helpfulness.

² There is empirical evidence suggesting that the relationship between numerical rating and review helpfulness is also nonlinear (see Mudambi and Schuff 2010). As a robustness check, we entered the squared term of ratings in these models, and all the results discussed below still holds.

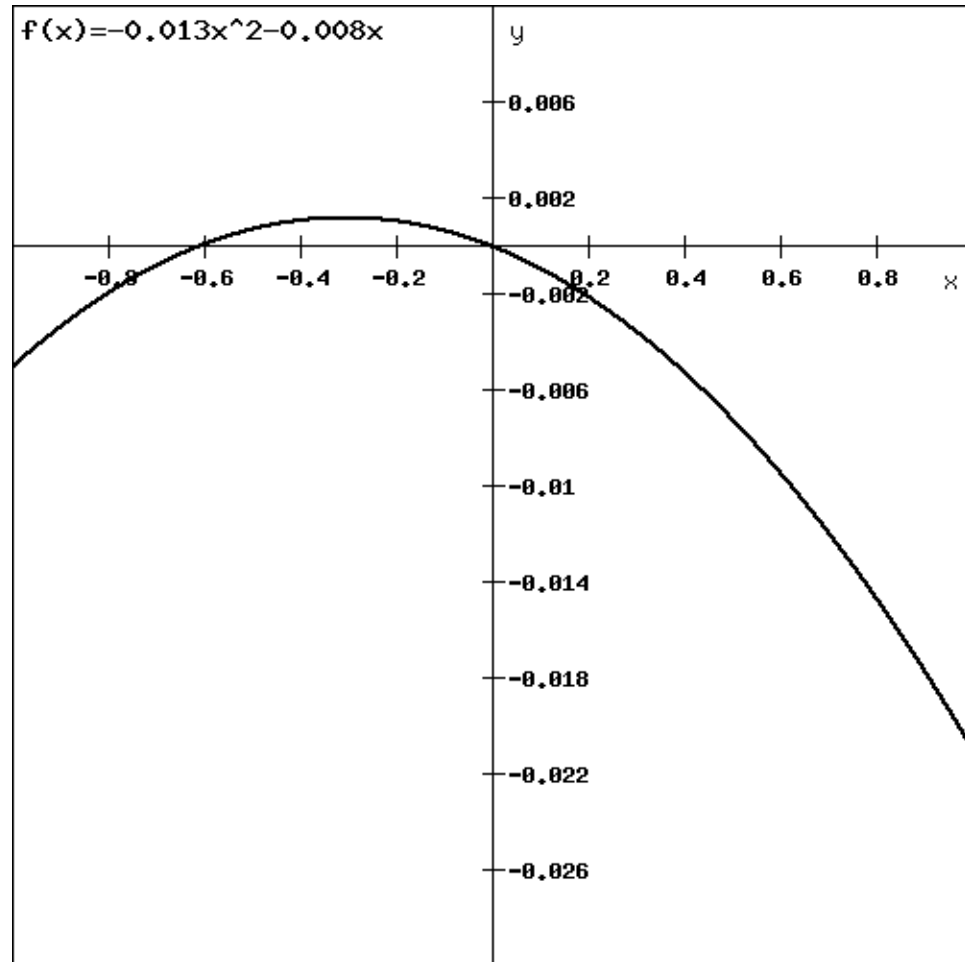


Figure 7: An Approximate Plot of the Non-linear Relationship Between Arousal and Review Helpfulness

3.3.4 Discussion

This study used actual review data from Apple’s App Store to provide empirical evidence for the first hypothesis. In the context of real app reviews, the relationship between arousal and review helpfulness is inverse U-shaped such that a higher level of arousal results in a less helpful review only after a certain level of arousal is achieved.

Notably, we utilized the percentage of people who casted a helpful vote for a review as proxy for review helpfulness, but this measurement might not be perfect. Two processes may underlie the effects we observed. First, the arousal level of reviewers may

influence the actual quality of reviews produced. This interpretation is compatible with our theory that arousal levels are associated with performance following an inverse U-shaped curve. Second, the arousal cues embedded in text reviews can also affect review readers and their perceptions. Because of people's lay theories about emotions, cues of various arousal levels may influence review readers' interpretations of reviews above and beyond their objective quality. We turn our focus to this possibility in the next section.

3.3.5 Lay Theories of Emotional Arousal

Next, we explore individuals' interpretation of emotional arousal through experimental methods. In particular, we focus on moderate to high levels of emotional arousal for two reasons. First, the content produced by reviewers with very low and very high levels of arousal tends to differ significantly, so it is infeasible to manipulate arousal across its full spectrum while keeping the substantive content identical. In addition, because reviewers are expected to care about an app, reviews often contain at least moderate levels of arousal. Reviewers who are not aroused at all may decide not to write a review in the first place. Hence, it is most relevant to examine the effect of arousal cues beyond moderate levels.

Given our focus on interpretations of emotional arousal, it is necessary to differentiate perceivers from actors. Review readers utilize review writers' conveyed emotions as a source of information to inform their evaluation of review helpfulness (see Van Kleef 2010). There is strong evidence that individuals notice and recognize emotional cues, including emotional words in verbal communication (Zeelenberg et al. 2006). In addition, literature examining the effect of arousal on attention to emotional

information suggests that perceivers readily attend to and utilize arousal cues (Kousta et al. 2009; Yiend 2010). Therefore, readers are likely to identify the emotional cues concerning arousal contained in a review.

As perceivers, review readers will utilize their lay theories to make sense of the emotions conveyed in a review (Murphy and Medin 1985; Shaver et al. 1987).

Particularly, the knowledge that human beings are bounded in rationality may guide how emotional arousal is interpreted. Bounded rationality assumes that decision makers are bound to make satisfactory, rather than optimal, choices in complex situations due to their unavoidable cognitive constraints (e.g., limitations in cognitive capacity) (Simon 1997). Following prior arguments, under moderate to high arousal levels, increases in arousal disrupt cognitive processing and polarize opinions (Gorn et al. 2001; White et al. 1981). Therefore, arousal constitutes a source of bounded rationality (Kaufman 1999). A higher level of arousal tends to be associated with reduced rationality, and it is through this association that perceived helpfulness is impacted.

We further posit that this association is more likely to exist in our lay theories for negative than for positive emotions. The literature testing the relationship between arousal and performance has examined exclusively negative emotions (Fantino et al. 1970; Hanoch and Vitouch 2004). From an evolutionary perspective, negative emotions are more critical than positive emotions for survival purposes (Frijda 2000), and negative emotions often evoke greater arousal than positive emotions. Given the adaptive priority of negative information over positive information (Dijksterhuis and Aarts 2003), people's folk theory about emotional arousal might be more complete for negative emotions.

Accordingly, we believe that the detrimental effect of arousal is more pronounced for negative emotions, and we propose the following hypotheses:

Hypothesis 2: For negative reviews containing moderate to high levels of arousal, additional arousal is perceived to be detrimental for review helpfulness.

Hypothesis 3: Beyond moderate levels of arousal, perceived rationality mediates the impact of emotional arousal on perceived review helpfulness for negative reviews.

To test these two hypotheses, we conducted another study. Study 2 was a lab experiment in which we directly manipulated arousal at moderate and high levels. We also explored the process by which arousal influences perceived review helpfulness.

3.4 Study 2: Experiment

In the laboratory experiment, we manipulated emotional arousal while holding substantive content constant. We constrained our manipulation of arousal to moderate to high levels, and focused on testing H2 and H3. Although these hypotheses concern only negative reviews, we also included positive reviews for comparison purposes. 186 undergraduate students from a southern U.S. university participated in the study. In our scenario, participants considered purchasing a fictional app from the App Store, based on their evaluation of four user reviews. Two of the user reviews (one positive and one negative) were treatment reviews containing cues signaling either high or moderate arousal. By comparing perceived review helpfulness across conditions, we were able to

examine the effect of emotional arousal. We also tested the mediating effect of perceived rationality by measuring this process variable.

3.4.1 Stimulus Materials

The app to be evaluated, “Fish Pond in Pocket,” was described as an interactive and realistic simulation of a pond, in which fish move across the screen in response to user input. Similar apps are available in the “entertainment” category in App Store (but none has the same name). This app was deemed adequate for our purposes because it appeals to general users and could reasonably result in either highly or moderately aroused reviews.

The stimuli used for the study were developed in two steps. First, we modified existing reviews retrieved from similar apps in the App Store, and created two reviews that were similar in content. We limited their length to one sentence, in order to strengthen the arousal manipulation in the next step. The negative review stated the following: “Who wants to watch fish move around on your screen and then swim off when you make a little movement?” and the positive stated the following: “I can touch the screen and see the fish swim away from my finger, as if I am actually there in water.” Second, we identified two pairs of negative emotional words and two pairs of positive ones; the two words in each pair differed in arousal. Based on prior research on emotional knowledge (Shaver et al. 1987), we selected the following word pairs: unhappy/outraged, disappointed/disgusted, glad/delighted, and happy/thrilled.

3.4.2 Procedure

Participants were asked to imagine that they were considering the purchase of a recently released, paid app, “Fish Pond in Pocket.” A description of the app was provided, as illustrated in Figure 8. The cover story explained that the researchers were working with the app developer to help improve their app quality and user experience, and as a result, participants would be evaluating a number of real reviews collected from Apple’s App Store.

Fish Pond in Pocket offers a realistic depiction of a pond with fish swimming around, casting shadows and changing course as real fish would. And it’s also interactive.

Fish Pond in Pocket Features:

- Touch Reactive
- Dynamic Water Simulation
- Virtual Fish
- Display Options

Figure 8: Description Screenshot of the “Fish Pond in Pocket” App

Participants read and evaluated four text reviews of the app, one at a time. Two positive ‘filler’ reviews were presented in positions 1 and 3 (e.g., “This app is one of the BEST for its price. My kids & friends love it.”). The two treatment reviews were presented in positions 2 and 4 of the sequence.

Participants were randomly assigned to either moderate arousal or high arousal conditions for the two treatment reviews. One treatment review was negative and the

other was positive, with their sequence counterbalanced. To manipulate emotional arousal, we appended emotional sentences both before and after the content sentence. The moderate-arousal and high-arousal emotional sentences were held as similar as possible (see Table 10).

Table 10: Treatment Reviews for App Experiment

Valence	Arousal	1st Sentence	2nd Sentence	3rd Sentence
Negative	Moderate	I'm unhappy that I paid for this app.	Who wants to watch fish move around on your screen and then	Disappointed with it.
	High	I'm outraged that I paid for this app!	swim off when you make a little movement?	Disgusted with it!
Positive	Moderate	I'm glad that I bought this app.	I can touch the screen and see the fish swim away from my finger,	Happy with it.
	High	I'm delighted that I bought this app!	as if I am actually there in water.	Thrilled with it!

After reading each treatment review, participants reported their perceptions of: (1) the helpfulness of the review, and (2) the perceived rationality of the reviewer. Perceived review helpfulness was measured on a 9-point semantic differential scale, using three items adapted from Sen and Lerman (2007). Perceived rationality was measured on a 9-point scale using two items adapted from Tormala, Clarkson and Henderson (2011). In addition to these dependent measures, participants also answered manipulation check questions with regard to the valence and arousal of the treatment reviews. These measures are presented in Appendix B.

3.4.3 Results

Analyses of manipulation check items revealed that reviews in the high arousal conditions were more aroused than those in moderate arousal conditions: negative ($M = 7.62$ vs. 5.51 , $p < 0.001$) and positive ($M = 6.51$ vs. 5.36 , $p < 0.001$). Therefore, the manipulation of emotional arousal was successful. Additionally, an analysis of perceived review valence showed that the high arousal negative review was more unpleasant than the moderate arousal negative review ($M = 1.55$ vs. 2.14 , $p < 0.001$), and the high arousal positive review was rated more pleasant than the moderate arousal positive review ($M = 7.96$ vs. 7.27 , $p < 0.001$). Given that high arousal polarized valence perceptions, we ran the following analyses for negative and positive reviews separately, and controlled for the perceived valence of the review.

Further analyses examined the reliability and validity of major constructs in the study. First, measures for perceived helpfulness and rationality showed adequate reliability (Cronbach's alphas > 0.80) (Nunnally 1967). Next, an exploratory factor analysis (EFA, principle components method with Varimax rotation) indicated adequate convergent and discriminant validity of the two constructs. For both treatment reviews, EFA consistently provided two factors. In addition, loadings of items on their respective factor were higher than 0.7, higher than the loadings of these items on the other factor (< 0.5), and higher than loadings of other items on this factor (Straub 1989).

Next, we examined the effect of arousal on perceived review helpfulness. First, ANCOVA was conducted for the negative review, with emotional arousal entered as a between-subject factor. We controlled for the perceived valence of the review and the sequence in which it appeared. In line with H2, pairwise comparisons revealed that the

difference in perceived helpfulness between high and moderate arousal conditions was significant ($M = 5.62$ vs. 6.20 , $p < 0.05$). Thus, despite having the same substantive content, a negative review that contained highly aroused emotions was considered less helpful than one that contained moderately aroused emotions. A similar analysis was performed for the positive review, but the difference in perceived helpfulness was not determined to be significant ($p > 0.5$). We will discuss about this finding later.

Next, we explored whether the effect of emotional arousal observed for negative reviews was mediated by perceived rationality. When perceived rationality was entered as the dependent variable, an ANCOVA showed results similar to those above. Specifically, the difference in review helpfulness between high and moderate arousal conditions was significant ($M = 4.66$ vs. 5.51 , $p < 0.01$). Further analysis revealed a significant initial relationship between emotional arousal and review helpfulness ($\beta = -0.58$, $p < 0.05$), which became non-significant after controlling for perceived rationality ($\beta = -0.18$, $p > 0.4$). The mediation was significant using Sobel's test statistic ($t = -2.60$, $p < 0.01$). Together, these findings indicate that the impact of arousal on the perceived helpfulness of a negative review can be explained by perceptions of reviewer rationality.

3.4.4 Discussion

By directly manipulating the emotional arousal of reviews and measuring their perceived rationality, Study 2 provided evidence supporting both H2 and H3. As predicted, participants considered highly aroused negative reviews to be less helpful than moderately aroused negative reviews, even when the objective content produced by

reviewers were held constant. Furthermore, this difference was mediated by perceived rationality.

However, we did not observe any evidence for the case of positive reviews. As explained earlier, although the high-arousal positive review was considered more aroused than the moderate-arousal positive review, this difference was not as dramatic as the one obtained for negative reviews. In other words, we were unable to increase the arousal level of the positive review as much as that of the negative review, even when we used highly aroused emotional words. This issue may arise from people's asymmetric processing of positive and negative emotions, and cannot be easily addressed using experimental methods.

3.5 General Discussions

To explore the effect of emotional arousal on review helpfulness, we conducted two studies and provided generally converging evidence for our hypotheses. Study 1 collected secondary review data from Apple's App Store and coded emotional arousal using software tools. As predicted by H1, results provided evidence that the relationship between review arousal and perceived helpfulness follows an inverse U-shaped curve. Extending these findings, Study 2 focused on interpretations of emotional arousal above and beyond the effects of objective review content, utilized an experimental method, and manipulated arousal directly. We found that above moderate levels, arousal had a damaging effect on helpfulness perceptions of negative reviews, and that this effect was mediated by perceived rationality. However, we did not observe similar effects for positive reviews.

3.5.1 Theoretical Implications

Existing research examining determinants of review helpfulness has focused mostly on easily observable variables, including ratings, reviewer location, etc (Chevalier and Mayzlin 2006; Forman et al. 2008; Mudambi and Schuff 2010). While recent efforts have begun to explore text content and show its nontrivial influence (Cao et al. 2011), our understanding of what constitutes a helpful review in terms of content is limited. In particular, emotional cues are widespread in textual reviews (Hennig-Thurau et al. 2004), and they are likely to affect how readers evaluate and make sense of the peer opinions expressed (Kuan et al. 2011). This paper is among the first attempts to disentangle the effect of emotion on review helpfulness, and thus contributes to the broader literature on affective processes in consumer decision-making.

In classic emotion frameworks, emotions are characterized along two fundamental dimensions: valence and arousal (Niedenthal 2008; Russell 1980). While research examining the valence (positive or negative) dimension abounds, fewer scholars have focused on arousal. Experimental studies manipulating arousal have shown effects on advertising effectiveness (Gorn et al. 2001; Sanbonmatsu and Kardes 1988), resistance to temptation (Fedorikhin and Patrick 2010), and even height perception (Stefanucci and Storbeck 2009). Arousal has also been found to increase social transmission of online content (Berger 2011; Berger and Milkman 2012). Our paper extended this literature and explored the non-linear effect of embedded arousal on information diagnosticity, using actual user reviews from Apple. These resulted in a more robust demonstration of the role of arousal in online word-of-mouth.

Moreover, we built on the attention-narrowing account of emotional arousal by proposing that the effect of emotional arousal proposed in H2 could be explained by perceptions of reviewer rationality. Study 2 provided evidence for this mediation in the case of negative (but not positive) reviews. More research may be needed to further explore the potential interaction between arousal and valence. However, the arguments for our hypotheses are general enough that they could reasonably be generalized to broader word-of-mouth settings (such as online communities, social media, etc.).

3.5.2 Practical Implications

Our findings not only deepen understanding of the role of emotional cues in text reviews, but also have important implications for reviewers individually and for review sites. For example, reviewers who want to leave negative reviews may be ‘fired up’ at the time, and, if so, our studies suggest that their reviews will appear less helpful to readers due to heightened perception of reviewer irrationality. To counter this effect, aroused reviewers might think more carefully in drafting their reviews and strive to describe their experience more accurately. Alternatively, they might focus on describing the objective experience while not conveying their emotions. At the other end of the spectrum, a reviewer who is too ‘relaxed’ may generate less helpful reviews as well. Users who are not aroused at all should spend more effort drafting their reviews. Apple currently does not have guidelines in place to guide app users through the review process. Our studies suggest that review sites stand to gain by designing and implementing corresponding writing guidelines. For example, they could require a minimum length of reviews to assure that users will spend adequate effort. Moreover, when a review is predicted to be

negative (e.g., based on star rating), the review sites can remind reviewers to think more ‘rationally’ and avoid ‘ranting without reasoning.’

App developers may benefit from our findings by learning how to better manage and respond to online reviews. Intuition might suggest that highly aroused, negative user reviews will be particularly harmful, but our studies indicate that higher arousal actually hurts the helpfulness of these reviews, due to a heightened perception of irrationality. On the other hand, reviews that lack emotional arousal may exert little influence over readers as well. Therefore, if targeting and responding to influential user reviews is a concern, app developers would be advised to pay particular attention to those indicating moderate levels of arousal.

3.5.3 Limitations and Future Research

This paper has several limitations that warrant future research. First, despite our focus on emotional arousal, we manipulated valence in Study 2 and controlled for it in both studies. Findings regarding an interaction between valence and arousal were not conclusive. The experiment was successful in manipulating arousal for positive reviews, but this manipulation was weaker than for negative reviews. Future studies could explore whether (and if so, why) the effect of arousal is contingent on the emotional valence.

Second, we proposed perceived rationality as a mediator to explain the effect of emotional arousal, but alternative explanations exist. Most notably, high arousal may be associated with reviewers that are objectively less helpful reviews in the first place. This explanation is consistent with our results, but we cannot tease this possibility out in our empirical study. Although Study 2 provided evidence that high arousal could reduce

review helpfulness even when the objective content is held constant, more research is needed to separate the two processes and explore their relative impact.

Finally, although this study was conducted in the context of Apple's App Store, we believe that the theoretical foundation and arguments can be generalized to other word-of-mouth settings. However, more research is needed to test the robustness of our findings in other contexts and to identify any contingencies of the effect of emotional arousal. Emotional communication is ubiquitous in the online environment, and how emotions communicated through online channels are perceived is a topic that warrants future scholarly attention.

3.6 Conclusion

Above and beyond the effect of ratings, objective information, and other factors, we believe that the emotions embedded in text reviews can powerfully impact consumers' evaluation of reviews. Two studies utilizing diverse methods provided largely converging evidence that heightened arousal leads to lower evaluations of review's helpfulness, but only beyond a certain optimal level of arousal. This paper extends our understanding of emotions in online word-of-mouth and highlights the importance of incorporating emotional arousal in consumers' judgment and decision-making processes.

CHAPTER 4

DISCRETE EMOTIONS IN ONLINE SELLER REVIEWS: AN EXPERIMENTAL AND EMPIRICAL EXPLORATION OF REVIEW HELPLEFULNESS

4.1 Introduction

By allowing prospective customers to consult peer evaluations when making their purchase decisions, online reviews have played an increasingly important role in the popularity and success of electronic commerce. Like other forms of online word-of-mouth, reviews help reduce uncertainty surrounding the shopping experience and inform future consumers (Dellarocas 2003). Specifically, the availability of opinions from previous customers has been shown to enhance buyer trust (Ba and Pavlou 2002) and increase sales (Chevalier and Mayzlin 2006; Forman et al. 2008). For purposes of this paper, ‘online reviews’ refer to peer-generated evaluations posted on company or third party websites (Mudambi and Schuff 2010). Although online reviews can be targeted towards either products or sellers, we focus on the latter, seller reviews, which have received surprisingly limited scholarly attention (e.g., Ba and Pavlou 2002; Pavlou and Dimoka 2006; Qu et al. 2008).

Our research concerns determinants of the perceived helpfulness of seller reviews. Following Mudambi and Schuff’s (2010) definition of a helpful product review, we define perceived helpfulness as the extent to which a peer-generated seller evaluation is perceived by consumers to facilitate their purchase decision process. Online retailers and third party sites often provide voting mechanisms to identify those reviews that are most

useful for assisting consumers in their purchase decisions. In the domain of product reviews, many online retailer interfaces (such as that of Amazon) sort user reviews according to their helpfulness ratings by default. In the domain of seller reviews, many e-commerce platforms have begun to offer this functionality as well (e.g., Yahoo! Shopping Merchant Reviews, ResellerRatings, Google Checkout Reviews). For instance, after each review in Yahoo! Shopping, the readers are presented with the question “Was this review helpful?” and the average rating assigned by prior readers is displayed (e.g., “4 out of 7 found this review helpful”).

In principle, these helpfulness voting systems assist prospective consumers seeking to make an informed decision while minimizing effort. Online reviews are unique from other forms of word-of-mouth in that a vast number of reviews are often available, and their authors are unknown. The availability of hundreds of reviews for a single seller may in theory provide more information to customers, but it also may create problems such as information overload (Jones et al. 2004). The sheer amount of reviews with varying content and quality makes it impossible for customers to evaluate and comprehend all the information before making a purchase decision (Liu et al. 2008). In many cases, consumers may only require a small set of helpful reviews, and the possibility of sorting based on helpfulness enables them to shorten information search, evaluate alternatives more efficiently, and make better purchase decisions (Cao et al. 2011; Mudambi and Schuff 2010).

A better understanding of perceived review helpfulness offers clear benefits to online retailers and review providers. The evidence indicates that helpful reviews are weighted more heavily when prospective customers make purchase decisions (Chen et al.

2008). If the most helpful reviews of an online retailer are also generally positive, the retailer can expect benefits in terms of reputation, trust, and sales. In contrast, if the most helpful reviews are largely critical, the retailer is likely to suffer. Although retailers are able to utilize voting mechanisms to identify helpful reviews, the accumulation of votes takes time (Zhang and Tran 2010). If more helpful reviews (especially negative ones) could be identified early on, even before votes have accumulated, then retailers could receive more immediate feedback and react accordingly. In addition, third-party review providers themselves stand to gain by providing high-quality reviews that bring potential value to both customers and retailers. Customers are more likely to visit these providers if doing so helps them to mitigate uncertainty and risk while making better decisions more efficiently (Dabholkar 2006). Websites that provide more helpful information than competitors stand to gain a strategic advantage in attracting attention and “stickiness” (Connors et al. 2011). In one prominent example, it is estimated that Amazon added \$2.7 billion to annual revenues by asking the simple question “Was this review helpful to you?” and using its interface to promote the most helpful reviews (Spool 2009). Meanwhile, customers who wish to leave seller reviews for the benefit of future shoppers may have little awareness regarding what constitutes a helpful review. Given a better understanding of the determinants of review helpfulness, websites may develop writing guidelines to encourage more useful seller reviews.

Scholars investigating review helpfulness have focused on a number of determinants that are easily observable, such as ratings and reviewer characteristics (Chevalier and Mayzlin 2006; Forman et al. 2008; Mudambi and Schuff 2010). Some scholars have also investigated the content and substance of reviews themselves (Cao et

al. 2011), finding that both the ratings and textual content of reviews can influence their helpfulness, and that extreme opinions tend to matter more in general. A common finding across these studies is a negativity bias, whereby negative reviews tend to be more influential than positive ones. Notably, however, the variables examined in all these studies have been non-emotional (e.g., numerical ratings, semantic characteristics of text reviews). In particular, prior work has tended to regard ‘negativity’ as a global construct, and no research has examined the distinctive roles of various negative emotions contained in negative reviews.

In recent years, researchers in the information systems field have started to investigate the role of emotions in various contexts. For example, IT adoption scholars have incorporated perceived affective quality (Zhang and Li 2005), perceived enjoyment (Sun and Zhang 2006b; Yi and Hwang 2003), and computer anxiety (Venkatesh 2000) into the user technology acceptance framework. Researchers examining web interface design have observed effects of users’ initial affective responses on subsequent behaviors (Deng and Poole 2010). In the field of trust and e-commerce, scholars have demonstrated the mediating effects of website users’ emotions on the development of online trust (Hwang and Kim 2007). We contribute to this emerging perspective, first by presenting a framework for the examination of specific emotions, and second by applying our framework to the context of seller reviews.

Extending earlier work on review helpfulness, we suggest that the emotions embedded in online reviews may have a crucial impact on their perceived helpfulness. Reviewers often express their feelings freely in textual comments, especially when the shopping experience is negative (Hennig-Thurau et al. 2004). When new customers read

these reviews, they are likely to perceive that emotional content is useful for understanding the reviews and making better decisions (Cao et al. 2011; Kuan et al. 2011). Extending the conventional wisdom of negativity bias, reviews with more negative emotions might logically be considered more helpful. However, in contrast to overall ratings, emotions are highly varied and complex, and cannot be reduced to simple ‘positive’ and ‘negative’ (Lerner and Keltner 2000). In particular, numerous different types of negative emotions occur in online reviews (anger, anxiety, disgust, etc.). Because these emotions are interpreted differently by readers, the effects of even same-valenced emotions may differ in systematic ways (Fontaine et al. 2007; Levenson 1992). In other words, emotions of similar valence may have distinct impacts on review helpfulness.

In order to address these issues, we ask the following research questions: How do reviewer emotions influence the perceived helpfulness of seller reviews? Specifically, does the impact of distinct emotions (such as anxiety and anger) differ in systematic ways, and what underlying mechanisms can be advanced to explain the differences? Drawing from research in judgment and emotion, we propose that the effects of specific negative emotions will vary due to underlying perceptions of reviewers’ cognitive effort. In particular, we argue that holding constant the objective information provided, anxiety-embedded reviews are considered more helpful than anger-embedded reviews, because anxious reviewers are assumed to think more carefully about the content they provide. These hypotheses are tested utilizing both experimental methods and a field study with archival data. Given that both methods provide converging evidence for our hypotheses, this methodological triangulation enables us to be more confident in our findings.

4.2 Literature Review and Hypotheses

4.2.1 Information Diagnosticity and Affect

As described above, the helpfulness ascribed to a review directly reflects the diagnosticity of the information that it contains. Information diagnosticity (Feldman and Lynch 1988) has been defined as “the degree to which one piece of information implies or determines one’s response to a given question or other circumstance requiring a judgment or behavior” (Feldman 1999, p. 48). In other words, a piece of information is considered diagnostic if it is informative for judgment and decision-making. This conceptualization of diagnosticity aligns with prior research suggesting that seller reviews are perceived as helpful to the extent that they are diagnostic for evaluating a seller (Kuan et al. 2011; Mudambi and Schuff 2010).

The diagnosticity concept provides an especially useful foundation for explaining negativity bias within online word-of-mouth. Abundant, cross-disciplinary evidence supports the existence of a generalized negativity bias, whereby “bad things will produce larger, more consistent, more multifaceted or more lasting effects than good things” (Baumeister et al. 2001, p. 325). Within e-commerce settings, negativity bias has been observed repeatedly in terms of the effect of ratings on review helpfulness (Cao et al. 2011; Kuan et al. 2011; Mudambi and Schuff 2010). One reason commonly cited for the bias is that negative information is less common than positive information, and consequently perceived as more diagnostic (Fiske 1980). Indeed, given that negative online feedback is much rarer than positive feedback (such as in eBay, see Resnick and

Zeckhauser 2002), it is not surprising that negative ratings would be considered more helpful by prospective customers.

Above and beyond the concept of negativity bias, we argue that the specific affective content embedded in online reviews plays a major role in determining their diagnosticity. Therefore, our interest lies in the consequences of affect-laden reviews. The term ‘affect’ describes a general category of mental processing that reflects subjective internal feelings (Cohen et al. 2008). Affective-based processing is typically contrasted with cognitive-based processing, although the relationship between affect and cognition has been a topic of great debate among psychologists. On one side of this debate, supporters of Zajonc (1980; 1984) argued that affect and cognition are separate and partially independent processes, and that affective responses can occur without cognition. In contrast, Lazarus (1982) and his supporters argued that the cognitive process of detecting and evaluating the meaning and significance of environmental stimuli (which need not be conscious) always precedes affective responses; i.e., affect requires cognition.

Subsequent authors have suggested that disagreement between these conclusions stems largely from different conceptualizations of ‘cognition’ and ‘cognitive process’ (Fulcher 2003). For the purpose of this research, we adopt a broader definition of these constructs. Therefore, we embrace the generally accepted position that cognition and affect are interdependent: not only can cognition precede affect, but affective states can also influence judgment and decision-making through cognitive processes (Forgas 1995). Most importantly, affective processing is often inherently cognitive, involving thoughts, judgments, and other cognitive elements, whether or not one is conscious of these

elements (Solomon 2008). As explained below, this is especially important for affective processing involving specific emotions, where cognitive factors can help differentiate diverse emotions (Scherer 2003) and explain affect-triggered inferential processes (Van Kleef 2010).

Because ‘moods’ and ‘emotions’ fall into the category of affective processes, it is helpful to distinguish the two terms (Lord and Kanfer 2002). Mood refers to a nonspecific, valenced feeling state that is typically low in arousal (Cohen et al. 2008); in contrast, emotion refers to “a mental state of readiness that arises from cognitive appraisals of events or thoughts” (Bagozzi et al. 1999, p. 184). For most affective researchers, an emotion differs from a ‘mood’ in that an emotion tends to be briefer, more intense, more context specific, and more intentional, with a particular cause (Ekman 1992; Frijda 1993). Emotions have a specific, known source, and may lead to specific coping actions. Furthermore, many emotions are directly coupled with specific resulting action tendencies and behaviors (Lerner and Keltner 2000). Although both mood and emotions undoubtedly play a role in the transmission of word-of-mouth, we focus on the emotion construct, because feelings expressed in seller reviews are targeted at particular purchase experiences and retailers.

4.2.2 Theories of Emotion

Emotions have been a subject of study across numerous disciplines, and as a result have been conceptualized in a variety of ways (Brosch et al. 2010). Among psychologists, two prominent approaches have been proposed for characterizing emotions. The first of these consists of dimensional theories of emotion, which assume

that all emotions can be shown to vary along a limited number of fundamental, abstract dimensions (Mano 1991; Watson and Tellegen 1985). Although no agreement exists regarding the ‘optimal’ number or naming of these dimensions (Larsen and Diener 1992; Russell and Mehrabian 1977), two or three dimensions have consistently emerged: valence (or pleasantness, evaluation), arousal (or activation, activity), and power (or potency, dominance). Among these three dimensions, valence is almost universally accepted, and there is evidence suggesting that valence and arousal are stable within and across cultures (Russell et al. 1989). A well-known example of this framework is Russell’s (1980) circumplex model, which maps the universe of emotion in a two-dimensional space along valence (“pleasant” vs. “unpleasant”) and arousal (“activated” vs. “deactivated”) (Niedenthal 2008). As a result, the model clearly delineates positive and negative emotions of different intensities. For example, a shopper presented with product recommendations may be seen to experience relief; the dimensional theories would explain this as a combination of positive valence and low arousal.

However, the dimensional view of emotions has increasingly been challenged (Smith and Ellsworth 1985). An oft-cited weakness of this approach is that global dimensions such as valence and arousal are less useful for capturing emotions that differ little across these fundamental dimensions (Fontaine et al. 2007). For instance, although fear and anger are very close to each other in the two-dimensional space characterized by valence and arousal (i.e., both emotions are unpleasant and activated) (Russell and Barrett 1999), they involve distinct phenomenology and tend to induce very different behaviors (Larsen and Diener 1992). More generally, given the variation and complexity

of emotional experience, other differences may have nontrivial influence on the way that they are experienced and resolved.

The other prominent approach consists of cognitive appraisal theories of emotion, which focus on the nuanced cognitive underpinnings of distinct emotional states (e.g., Scherer et al. 2001; Smith and Ellsworth 1985). This approach argues that emotional reactions to an event are a direct result of personal interpretations (appraisals) of the event itself and the situational environment (Frijda 1986; Roseman 1984). Therefore, emotions can be differentiated by a set of standard appraisal criteria (Ellsworth and Scherer 2003); these theories maintain that each distinct emotion is elicited by a unique pattern of cognitive appraisals, and that situations with the same appraisal pattern will induce the same emotion (Roseman and Smith 2001). For example, a shopper presented with product recommendations might appraise the event in terms of the unexpectedness of the advice (leading to surprise), the reduction in required effort (leading to relief), or the loss of personal control (resulting in anger), among other possibilities.

A large body of work has focused on identifying parsimonious sets of appraisal dimensions (Roseman 1984; Smith and Ellsworth 1985). The resulting frameworks vary considerably but contain a number of common appraisal dimensions, including pleasantness, certainty, and control. Pleasantness describes the extent to which one feels the situation is conducive to one's goals, certainty describes the extent to which events are predictable vs. unpredictable, and control describes the extent to which events are brought about by individual agency vs. situational agency (Smith and Ellsworth 1985). The emotions that we consider next, anxiety and anger, are both characterized by low pleasantness; however, they differ considerably in appraisals of certainty and control.

Anxiety arises from a situation that is appraised as unpredictable and dictated by the situation itself rather than by individuals; in contrast, anger arises from a situation that is appraised as predictable and dictated by other individuals (Lerner and Keltner 2000). Next, we elaborate on the implications of certainty appraisals in a review context, and how they may explain unique downstream effects of anxiety and anger.

4.2.3 Discrete Emotions and Cognitive Effort

Above and beyond their affective consequences, the appraisals that define an emotion often have carry-over effects on judgment and behavior (Ellsworth and Scherer 2003). Specifically, widespread evidence indicates that emotions create a predisposition to appraise subsequent events in line with the appraisal patterns characterizing those emotions (i.e., an “appraisal tendency”) (Lerner and Keltner 2000; Lerner and Keltner 2001). As a result, subsequent judgments and decisions will be interpreted in ways consistent with the underlying appraisals (e.g., Berkowitz and Harmon-Jones 2004; Keltner et al. 1993). For example, one feels sad when a negative event is appraised as situation-controlled (e.g., a natural disaster), because sadness arises from appraisals of situational control of negative events. Moreover, sadness often triggers a tendency to perceive situational control in subsequent, unrelated situations. Thus, sad people tend to blame the situation rather than specific parties for subsequent negative events.

For present purposes, an especially relevant argument among appraisal-tendency proponents is that emotions will have predictable effects on the level of subsequent cognitive effort, and that the key determinant of these effects is the underlying appraisal of ‘certainty’, i.e., the degree to which future events are predictable and comprehensible

(Lerner and Tiedens 2006). Table 11 provides a sample illustration of emotions that differ in valence and certainty appraisals (see Roseman 1984; Smith and Ellsworth 1985). Emotions associated with uncertainty (e.g., fear) have been shown to induce systematic, “mindful” processing that involves considerable thought and cognitive effort (Tiedens and Linton 2001). For instance, fearful people tend to process information more carefully and make efforts to reduce the uncertainty (see Lerner et al. 2003; Raghunathan and Pham 1999). On the other hand, emotions associated with appraisals of certainty (e.g., anger) have been shown to induce more “mindless,” heuristic processing that requires less directed thought and relies on automatic cues or “rules of thumb” (Bond et al. 2008; Chaiken and Trope 1999). For instance, people induced to feel angry tend to make punitive and hostile inferences, base their judgments on stereotypes, and pay less attention to the argument quality (see Bodenhausen et al. 1994; Lerner et al. 1998; Tiedens 2001; Tiedens and Linton 2001). In sum, when compared to certainty-appraised emotions, uncertainty-appraised emotions elicit a carryover effect of greater subsequent cognitive effort, even on unrelated tasks.

Table 11: Sample Emotions Characterized by Valence x Uncertainty

		Valence	
		Positive	Negative
Certainty	Low	Surprise	Fear
	High	Happiness	Anger

Our discussion thus far illustrates the advantage of an emotion-based approach to review content over an approach based on valence alone. To the extent that judgment depends on the appraisals associated with an emotion rather than its valence, two

‘positive’ emotions may affect decisions quite differently, while a ‘positive’ and ‘negative’ emotion may have similar effects. Therefore, experimental researchers utilizing an appraisal approach typically compare emotions that: 1) are differentiated by certain appraisal dimensions, 2) on judgments that relate to those dimensions (Han et al. 2007; Lerner and Tiedens 2006). In keeping with this approach, to illustrate the distinct effect of various emotions on perceived cognitive effort and review helpfulness, we need to select emotions that vary substantially in the appraisal dimension of certainty.

The two emotions this research focuses on are anxiety and anger. Although definitions for these two emotions can vary (see above), we adopt the adaptive/functional approach suggested by Lazarus (1991). Thus, we define ‘anxiety’ as an emotional state that motivates a person to avoid potential harm arising from ambiguous threat, and ‘anger’ as an emotional state that motivates a person to alleviate personal harm attributed to others. Given that both emotions represent one’s response to threat, anxiety and anger appear to share certain neurological underpinnings and are sometimes linked in psychiatric and clinical discussion (Danesh 1977; Rothenberg 1971). However, evidence for their phenomenological and functional independence is robust in the emotional literature, and they are typically treated as distinct (Oatley and Johnson-laird 1987).

The emotions of anger and anxiety are especially relevant to e-commerce settings, because both are commonly encountered in seller reviews (see later for specific examples). Anxiety in a review often stems from ambiguity regarding product quality, shipment times, or refunds/returns, while anger often results from mishandled transactions, inadequate customer service, or poor product performance.

The more important reason we focus on anxiety and anger is that they are similar in appraisals of pleasantness but differ heavily in appraisals of certainty. Although these two emotions also differ strongly in the control dimension, it is the appraisal of certainty that is the most relevant for our purpose of demonstrating the distinct impact of diverse emotions; we will discuss this issue later. Due to their distinction in certainty appraisals, therefore, anxiety and anger should have distinct effects on the effort expended by a reviewer. Specifically, anxiety results in increased sensitivity to threat and heightened vigilance to situational stimuli, and hence should enhance effort, whereas anger prompts reliance on superficial cues and more stereotypic thoughts, thus decreasing effort (Bodenhausen et al. 1994; Tiedens 2001).

4.2.4 Representation of Emotion Concepts

Because our hypotheses concern the downstream effects of reviewer emotions, we must differentiate perceivers (i.e., review readers) from actors (i.e., review writers). Our primary argument is that review readers use review writers' emotional expressions as a source of information to inform their judgment of review helpfulness (see Van Kleef 2010). In other words, a reader's assessment of a writer's cognitive effort is driven in part by emotions embedded in the review. For example, the logic above implies that anxious reviewers will tend to engage in more systematic processing than angry reviewers. Although this relationship may not always hold, our argument only requires that perceivers assume the relationship exists, based on their mental representation of emotion concepts.

To make this claim, we first argue that review readers are able to identify discrete emotions in the content of seller reviews. An abundance of evidence confirms that people attend to and recognize emotional cues. Historically, much of this research involved facial and bodily expressions (Atkinson et al. 2004; Ekman and Friesen 1971), but more recent work on verbal communication has shown that people not only recognize emotional words (Zeelenberg et al. 2006), but also distinguish between discrete emotions embedded in writing (Lindquist et al. 2006). Therefore, to the extent that a review contains emotional content, readers are likely to identify that content.

Next, we argue that readers who have identified emotional content in a review will make appraisal-consistent inferences about the reviewers' cognitive effort. Emotional displays perform an important social communication function, allowing observers to infer the sender's internal state (Keltner and Haidt 1999; Van Kleef 2009). Individuals' naïve, theory-based representations of emotion concepts will determine how they interpret and evaluate the emotions expressed by others (Murphy and Medin 1985; Shaver et al. 1987), and there is substantial evidence that these naïve theories accurately associate emotion categories with corresponding appraisals (Scherer and Grandjean 2008; Siemer and Reisenzein 2007). In an especially relevant example, Siemer (2008) demonstrated that participants' naïve theories of anxiety and anger aligned well with the tenets of appraisal theory. Therefore, emotions expressed in a review should inform readers about associated appraisals (Manstead and Fischer 2001); in our context, this means that readers are likely to associate anxiety with uncertainty and anger with certainty. Furthermore, lay theories of emotion include additional information about their behavioral consequences (Niedenthal 2008). Thus, due to the greater uncertainty

associated with anxiety, readers should perceive greater cognitive effort from anxious reviewers than angry reviewers.

Finally, it is reasonable to assume that readers will associate cognitive effort with review helpfulness; i.e., the more effort reviewers are perceived to have expended in sharing their experience, the more helpful the reviews are likely to be considered. Therefore, to summarize our arguments, reviews that contain content indicative of anxiety (vs. anger) will result in a higher level of perceived cognitive effort, which will in turn lead to perceptions that the reviews are more helpful. Our theoretical framework is illustrated in Figure 9. (Note that although we present perceived cognitive effort as a mechanism underlying the distinct impact of anxiety and anger, we do not preclude the possibility of other mechanisms.) Our hypotheses are as follows:

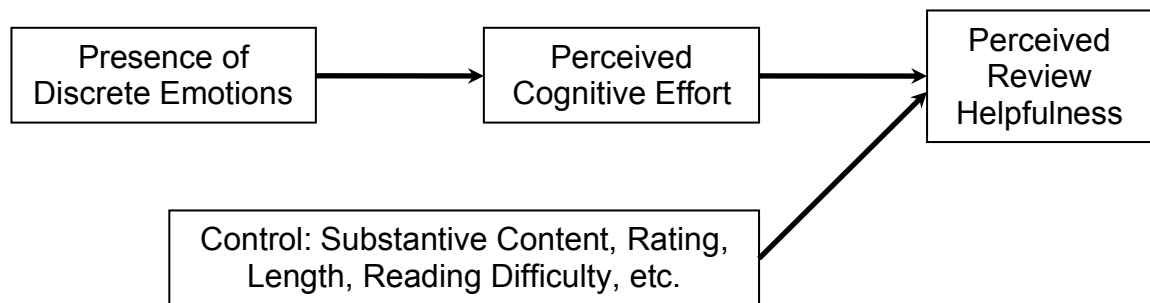


Figure 9: Theoretical Framework

Hypothesis 1: Anxiety-embedded reviews are perceived to be more helpful than anger-embedded reviews.

Hypothesis 2: Perceived cognitive effort mediates the differential impact of anxiety and anger on the perceived helpfulness of reviews.

In order to test these hypotheses, we conducted three studies utilizing distinct methods. In Study 1, we conducted a controlled experiment in which anxiety and anger were manipulated directly while controlling for potential differences in objective review content, and we explored the process by which these affect perceived helpfulness. In Study 2, we replicated the results of the first study while using a different manipulation to rule out alternative explanations. In Study 3, we extended the experimental results by examining actual seller reviews from a popular online platform (Yahoo! Shopping), in order to measure the impact of emotional content on ratings of review helpfulness.

4.3 Study 1: Experiment

In this study, we utilized a laboratory experiment to directly manipulate anxiety and anger in seller reviews, while controlling for potential differences in substantive content. Seventy-eight undergraduate students from a southern U.S. university participated in the study. Using a simulated seller feedback scenario, each participant was exposed to reviews of three potential stores: one written by an anxious reviewer, one by an angry reviewer, and one by a non-emotional reviewer. By comparing the perceived helpfulness across stores, we identified the differential impact of anxiety and anger on review helpfulness. In addition, we measured the process variable of perceived cognitive effort and tested its mediating effect.

4.3.1 Stimulus Materials

The preparation of stimuli for Study 1 involved two steps: (1) identification of text reviews lacking explicit emotional cues, so that substantive content would not

interfere with the emotion manipulation, and (2) appending an emotional sentence to the otherwise ‘non-emotional’ reviews, so that emotions could be varied directly. Therefore, we began with a set of actual seller reviews (see details in Study 3). From this set, we first selected a pool of one-star negative reviews, and then revised them by removing any sentences that directly indicated reviewer emotions. Of the revised pool, we selected 13 reviews containing content that could reasonably have been written by anxious or angry customers. Next, we conducted pretests to identify from reviews from the pool that were perceived to reflect equivalent levels of anxiety and anger: specifically, 25 pretest subjects were asked to read all the reviews in turn, and rate the anxiety and anger of each review author. Based on the results, we selected three reviews for the experiment, shown in Table 12. For each of the selected reviews the difference between anxiety and anger evaluations was not significant ($p > 0.8$).

Table 12: Review Stimuli Used in Experiment*

#	Text Review Content	Anxiety	Anger
1	I purchased a camera on Feb 27th for two day delivery and on March 23 I am still waiting for it, plus they billed me for it on Feb 27th.	6.7	6.8
2	Ordered a laptop battery (12 cell) and RAM. I received a 6 cell battery and the incorrect RAM. I returned the products to this merchant three weeks ago (and they were received), but still have not received my refund.	6.8	6.8
3	I placed an order on Dec. 14th using standard shipping because it said if I ordered by the 19th it would be delivered before Christmas. I just received an E-mail saying they shipped it today (Dec. 23rd) and estimated arrival date is Dec. 30th.	6.5	6.2

Notes:

* Perceived emotions of each review writer, measured by the question “In your opinion, to what extent does each of the following words describe how the reviewer felt when he/she wrote the above review?” Items included “anxious” and “angry,” and were measured on a 9-point scale (“not at all” to “very much”).

4.3.2 Procedure

As a cover story, participants were introduced to a fictitious third-party review site, “OnlineConsumerReview.com,” providing consumer reviews of online stores. The cover story explained that the researchers were working with this site to improve its data mining algorithms, and that to aid in this process, participants would be evaluating a series of real text reviews collected from the OnlineConsumerReview.com site.

Participants read and evaluated six text reviews, one at a time, each describing a different online store. The three treatment reviews were presented in positions 2, 4, and 6 of the sequence. Three ‘filler’ reviews were presented in positions 1, 3, and 5 of the sequence; these filler reviews were positive overall (e.g., “I liked their web site - lots of items with a decent description of each. Received exactly what I ordered in a timely manner ...”).

For the first treatment reviews, emotion was manipulated directly by varying the sentence appearing at the beginning of the review. In the *anxiety* condition, the review began with the sentence “My experience with this seller has caused a lot of anxiety.” In the *anger* condition, the review began with the sentence “I was very angry after everything that happened.” The review in the *baseline* (control) condition contained no additional upfront sentence. Note that due to the within-subject design of the study, substantive content of reviews across the three conditions could not be held identical without appearing artificial. Therefore, we fixed the sequence of treatment reviews constant, but counterbalanced the order in which the treatments occurred. In this way,

each of the three reviews appeared in each of the three conditions (anxiety, anger, baseline) an equivalent number of times.

After reading each review, participants reported their perceptions of: 1) the helpfulness of the review, and 2) the cognitive effort expended by the reviewer. Perceived review helpfulness was measured on a 9-point semantic differential scale, using three items adapted from Sen and Lerman (2007). Perceived cognitive effort was measured on a 9-point scale ranging from “not at all” to “very much”, using three items adapted from Huddy, Feldman and Cassese (2007). These measures are presented in Appendix C.

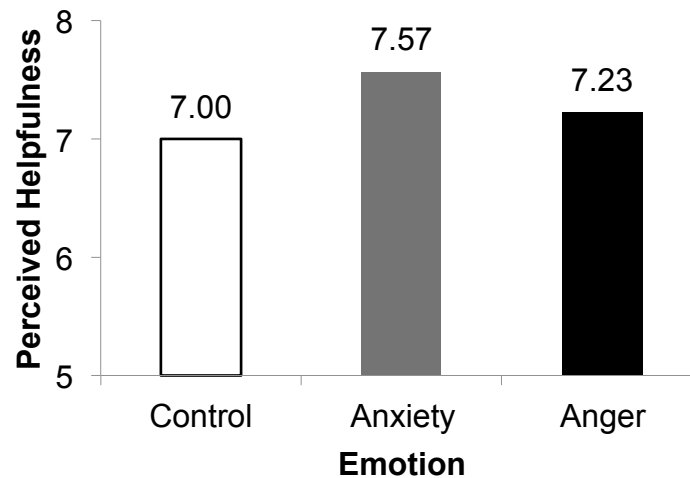
4.3.3 Results

Before further analysis, we conducted a manipulation check of the stimulus materials to ensure that emotional content was correctly identified. A separate group of 30 subjects underwent a procedure that was similar to the main study; however, the dependent measures after each review were replaced with the following question: “In your opinion, to what extent does each of the following words describe how the reviewer felt when he/she wrote the above review?” Response options included “anxious,” “angry,” “sad,” and “happy” (1 = “not at all” and 9 = “very much”). Analysis was performed using pairwise comparisons after a repeated-measure ANCOVA (controlling for the order of reviews). Confirming that the treatment reviews successfully targeted their relevant emotions, reviews in the anxiety condition were more related to anxiety than to anger ($M = 8.27$ vs. 7.20 , $p = 0.013$), and reviews in the anger condition were more related to anger than to anxiety ($M = 8.70$ vs. 6.27 , $p < 0.001$). Additionally,

reviews in the control condition were related to both anxiety and anger to a similar extent ($M = 6.87$ vs. 7.17 , $p = 0.344$).

We also examined the reliability and validity of major constructs in the study. For each of the three treatment reviews, Cronbach's alphas for both constructs were well above 0.80, demonstrating adequate internal consistency reliability (Nunnally 1967). Next, we conducted an exploratory factor analysis (EFA) to assess convergent and discriminant validity of the two constructs, utilizing the principle components method with Varimax rotation. For each review, EFA consistently provided two factors. Moreover, in the rotated component matrix: loadings of items on their corresponding factor were higher than 0.7, higher than loadings of other items on this factor, and higher than the loadings of these items on the other factor (< 0.5) (Straub 1989). Together, these results indicated adequate convergent and discriminant validity.

The first important question concerns whether perceived helpfulness varied across anxious vs. angry reviews. The pattern of means for perceived helpfulness is illustrated in Figure 10. A repeated-measure ANCOVA was performed to examine the difference in perceived helpfulness across treatment reviews. Emotional condition was entered as a within-subject factor, and the counterbalancing of the three treatment reviews was entered as a covariate. In line with H1, pairwise comparisons revealed that the difference in perceived helpfulness between anxiety and anger conditions was significant ($M = 7.57$ vs. 7.23 , $p < 0.05$). Thus, reviews containing anxiety were considered more helpful than those containing anger, despite having the same objective content.



Notes:

- The review in the anxiety condition began with the sentence “My experience with this seller has caused a lot of anxiety.”
- The review in the anger condition began with the sentence “I was very angry after everything that happened.”

Figure 10: Perceived Helpfulness of Seller Reviews Across Emotion Conditions

In a supplementary analysis, we compared the helpfulness of emotional reviews with that of the baseline review. Pairwise comparisons showed that anxious reviews were considered significantly more helpful than baseline reviews ($M = 7.57$ vs. 7.00 , $p < 0.001$), whereas angry reviews were not reliably different from baseline ($M = 7.23$ vs. 7.00 , $p = 0.16$). Taken together, these results indicate that negative reviews were considered more helpful if they indicated anxiety but not if they indicated anger.

Next, we explored whether the differential effects of anxiety and anger on perceived helpfulness were mediated by perceived cognitive effort. The pattern of means for perceived effort is illustrated in Figure 11. When perceived cognitive effort of reviewers was entered as the dependent variable, a repeated-measure ANCOVA showed results similar to those above. Specifically, the difference in review helpfulness between

anxiety and anger conditions was significant ($M = 6.27$ vs. 5.82 , $p < 0.01$); as was the difference between anxiety and control conditions ($M = 6.27$ vs. 5.46 , $p < 0.001$), and the difference between anger and control conditions ($p < 0.05$). In sum, anxious reviewers were perceived to have spent more effort than angry reviewers in writing the review.

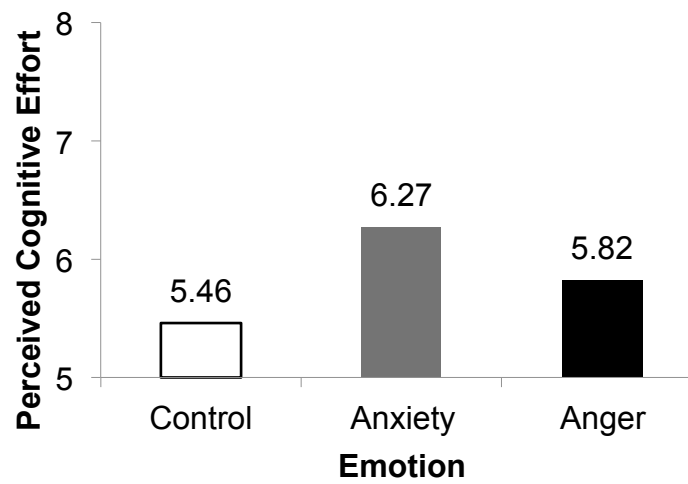


Figure 11: Perceived Cognitive Effort of Reviewers Across Emotion Conditions

Although various procedures exist for testing mediation, the most common methods (e.g., Baron and Kenny 1986) apply only to cases in which the treatment varies between (rather than within) participants. Therefore, we employed the two-criteria procedure proposed by Judd, Kenny and McClelland (2001) for testing mediation in within-subject designs. Results indicated that both criteria were satisfied: first, the difference in perceived cognitive effort between anxiety and anger reviews was significant ($M = 6.27$ vs. 5.82 , $p < 0.01$) and in the same direction as the difference in perceived helpfulness. Second, the difference in perceived cognitive effort between anxiety and anger reviews was predictive of the difference in perceived helpfulness ($p <$

0.001). Therefore, perceived cognitive effort was confirmed to mediate the differential impact of anxiety and anger on perceptions of review helpfulness.

4.3.4 Discussion

By directly manipulating discrete emotions and measuring perceived cognitive effort, Study 1 provided evidence supporting our hypotheses. As predicted by H1, participants considered anxious reviews to be more helpful than angry reviews, despite the fact that any meaningful information in the reviews was controlled for. Additionally, we showed that this difference could be explained by the perceived cognitive effort of the reviewer, providing evidence for H2.

Although results of the study supported both hypotheses, other explanations may be advanced to account for our results. Two such explanations concern the valence and arousal of the treatment reviews. In terms of valence, the existence of a generalized ‘negativity bias’ suggests that negative information is rarer and thus considered more diagnostic (Baumeister et al. 2001); if so, the anxious reviews may have been rated more helpful simply because they were more negative. In terms of arousal, ample evidence exists that high levels of arousal can impair executive function and induce “mindless” heuristic processing, characterized by low elaboration and effort (Eysenck 1982; Humphreys and Revelle 1984; Mueller 1979; Sanbonmatsu and Kardes 1988). Thus, compared to anxious reviews, the angry reviews may have been associated with less effort due to their higher levels of arousal. A different possibility concerns attributions of reviewers’ motivations (Sen and Lerman 2007): according to correspondent inference theory (Jones and Davis 1965), perceivers will tend to attribute an actor’s behavior to

stable dispositions rather than external reasons, unless the behavior is unusual or unexpected. Hence, if angry reviews are considered more typical than anxious reviews, they may invoke more dispositional attributions (e.g., “the author is easily irritated”), and thus be considered less helpful. Finally, it may be the case that participants felt more empathy towards the authors of the anxious reviews (Lazarus 1991). If so, they may have perceived the anxious reviews as more helpful based on their empathic response rather than inferences regarding reviewer effort.

To address these explanations directly, we conducted a follow-up study. Forty-nine undergraduate students underwent a procedure similar to that described above; however, the dependent measures were replaced by a series of questions addressing the valence, arousal, attributions, and empathy associated with each review. All questions were adapted from scales used in prior literature (see Appendix). Analyses were performed through a repeated-measure ANCOVA controlling for the order of reviews. Results indicated that contrary to a valence-based explanation, reviews in the anxiety conditions were rated (marginally) less negative than those in the anger conditions ($M = 1.92$ vs. 1.59 , $p = 0.072$). Contrary to explanations based on attributions or empathy with the reviewer, comparisons of anxiety and anger conditions showed no reliable differences in these measures ($p > 0.2$). On the other hand, an explanation based on arousal could not be ruled out, as reviews in the anxiety conditions were perceived to be lower in arousal than reviews in the anger conditions ($M = 6.50$ vs. 7.23 , $p = 0.018$). However, given that elevated arousal is a fundamental component of anger but not anxiety (Smith and Ellsworth 1985), this result may be consistent with our arguments, to the extent that the higher arousal of angry reviewers is associated with less cognitive effort. In summary, the

follow-up provided evidence contrary to various competing explanations; Study 2 investigated these issues further by incorporating an alternative design.

4.4 Study 2: Experiment

The primary goal of Study 2 was to replicate the main findings of the first study utilizing a between-subject design. This design made it possible to hold constant the substantive content of the review, strengthen the manipulation of emotion, and control for review valence and arousal in the data analysis.

4.4.1 Procedure

Seventy-three undergraduates participated in this study and were randomly assigned to either the *anxiety* or *anger* condition. The cover story and procedure was similar to that of Study 1, with three major exceptions. First, only one review was evaluated (review #2 in Table 12). Second, the emotion manipulation was strengthened by appending a sentence at both the beginning and end of the review. Specifically, the review began with the sentence “I feel so worried (mad) as I’m writing this!” and ended with the sentence “Let me tell you: I’m very nervous (irritated).” Finally, in addition to the dependent measures described in Study 1, participants also provided evaluations of valence, arousal, attribution, and empathy (see Appendix C). At the end of the procedure, they also completed the emotion manipulation check described above.

4.4.2 Results

Analyses of manipulation check items revealed that the review in the anxiety condition was more related to anxiety than to anger ($M = 7.86$ vs. 4.59 , $p < 0.001$), and the review in the anger condition was more related to anger than to anxiety ($M = 8.56$ vs. 4.89 , $p < 0.001$). Thus, the manipulation of emotion was successful.

Next, ANCOVA was performed to examine the perceived helpfulness of anxiety-embedded and anger-embedded reviews, while controlling for the effect of valence and arousal. Replicating the results of the first study, and in line with H1, perceived helpfulness was significantly higher in the anxiety condition than the anger condition ($M = 7.33$ vs. 6.26 , $p < 0.05$).

Next, we examined the evidence for mediation effects predicted by H2. Analyses revealed a significant initial relationship between emotion and review helpfulness ($\beta = -1.07$, $p = 0.020$) that became non-significant after controlling for perceived effort ($\beta = -0.19$, $p = 0.210$), indicating the presence of mediation (Sobel test statistic = -2.12 , $p = 0.034$). Consistent with Study 1 and H2, these findings indicate that the differential impact of anxiety and anger on the perceived helpfulness of a review is mediated by perceptions of the reviewer's cognitive effort.

Finally, we explored the alternative explanations discussed in Study 1. First, *t*-tests revealed that the review in the anxiety condition was considered less negative than the review in the anger condition ($M = 2.38$ vs. 1.41 , $p < 0.001$), ruling out the alternative explanation of valence. Second, arousal in the anxiety condition was considered to be lower than that in the anger condition ($M = 6.65$ vs. 8.32 , $p < 0.001$), suggesting that arousal may indeed play a role in the differential impact of the two emotions. Given that

the analyses above controlled for arousal, this argument cannot account for our findings; however, it is clearly worthy of future exploration. Third, attribution measures revealed that dispositional attributions were marginally greater for the anxiety review than the anger review ($M = 5.03$ vs. 3.97 , $p = 0.074$), contrary to an account based on attribution. Lastly, empathy measures were virtually identical across the anxiety and anger conditions ($p > 0.5$).

4.4.3 Discussion

Study 2 replicated and extended the findings of Study 1 in a between-subject design which held constant objective content and controlled for possible effects of valence and arousal. Findings provided evidence supportive of both our hypotheses, while ruling out a number of alternative explanations.

The main advantage of the experimental method utilized in Study 1-2 was the ability to manipulate anxiety and anger in a straightforward manner. This parsimony enabled us to avoid potential confounds and directly explore the reasons for differential effects of anxious and angry reviews. On the other hand, our designs required a degree of artificiality in both the experimental task and the reviews themselves. We address these concerns in Study 3 by examining real-world seller reviews.

4.5 Study 3: Yahoo Merchant Reviews

The primary goal of Study 3 was to test H1 and explore the effects of discrete emotions on review helpfulness in a real-world setting. To do so, we collected and analyzed actual review data from the Yahoo! Shopping website, which provides both user

ratings and text reviews for online merchants. At the time of data collection, the site had accumulated over eight years of customer reviews. In a review, prior consumers of a merchant can evaluate that merchant by leaving a rating on a scale of 1 to 5 stars. Additionally, they can write a text review to provide more details about their experience with the merchant, as illustrated in Figure 12. The review page of each merchant displays all reviews for that merchant chronologically, and the most recent reviews appear first by default.



Figure 12: Screenshot of a Yahoo Retailer Review

4.5.1 Data Collection

Individual reviews were used as the unit of analysis, and data collection took place in April 2011. Merchants are classified into various categories based on the products they sell (books, electronics, software, etc.). We began by targeting stores in the “electronics” category, which sell a wide range of products including cameras, cell phones, televisions, MP3 players, home video, etc. We retrieved all the historical reviews of each merchant; those without a single review were eliminated, leaving 167 stores. For each of these merchants, we collected the following information from each review:

rating, text review content, helpful votes, and total votes. We also collected store-level information, including the average rating and count of all ratings of each store.

In order to reduce noise in the reviews, the following steps were taken. First, 562 reviews which included non-ASCII characters (mostly from non-English languages) were removed. Next, we removed reviews that contained no text content (4,571), reviews that contained only “EOM” (“End of Message,” 27,708), and reviews that contained only symbols or dates (10). These steps resulted in 154,834 reviews. Of this group, only 7,322 reviews (4.7%) had received any helpfulness votes (see below), which is not unusual in online review settings (Pavlou and Dimoka 2006). Analysis was conducted on this set of 7,322 reviews.

4.5.2 Variables

The dependent variable of interest, review helpfulness, was operationalized as follows. Below each review, Yahoo! Shopping lists the question “Was this review helpful?”, along with “Yes” and “No” options. A review that has received at least one vote will display the number of “helpful” votes and total votes immediately before the review content. Helpfulness was measured as the proportion of “helpful” votes out of the total votes a review received (i.e., the number of people who voted “Yes” divided by the total number of people who cast a vote). Therefore, the value of helpfulness ranged from 0 to 1, with a higher percentage indicating a more helpful review. The average helpfulness of the analyzed reviews was 0.68, indicating that most reviews in the final set were considered relatively helpful. Tables 13 and 14 present a summary of statistics and correlations for this and the remaining variables (described below). Because no

reasonable measure of perceived cognitive effort was included in the Yahoo! merchant data, Hypothesis 2 could not be tested directly in this study.

Table 13: Descriptive Statistics for Final Review Pool (N = 7,322)*

Variable	Mean	Std. Dev.	Min	Max
1 Review helpfulness	0.68	0.40	0	1
2 Rating	3.29	1.81	1	5
3 Length	69.82	70.76	1	707
4 Reading difficulty	10.32	4.25	-10.2	121.5
5 Anxiety	0.17	1.00	0	50
6 Anger	0.19	1.14	0	50

Notes:

* Refer to Table 16 for the definitions of these variables.

Table 14: Variable Correlations for Final Review Pool (N = 7,322)

Variable	1	2	3	4	5	6
1 Review helpfulness	1					
2 Rating	-0.158	1				
3 Length	0.182	-0.376	1			
4 Reading difficulty	-0.051	0.226	-0.240	1		
5 Anxiety	0.006	-0.050	-0.006	0.068	1	
6 Anger	-0.002	-0.132	0.027	-0.018	0.018	1

Measurement of discrete emotions in the text reviews was conducted with Linguistic Inquiry and Word Count (LIWC), a text analysis software program developed by Pennebaker and colleagues (2007). LIWC was designed to efficiently evaluate psychological and structural components of text samples. The tool has been widely adopted in psychology and linguistics (Tausczik and Pennebaker 2010), and its reliability and validity have been investigated extensively (Pennebaker et al. 2007; Pennebaker and Francis 1996). LIWC includes a psychometrically validated internal dictionary comprised

of approximately 4500 words and word stems, each of which is classified into one or more categories. After receiving a text sample, the software processes each word in the sample, one at a time. As each word is processed, LIWC searches its dictionary file for a match, and if a match occurs, the appropriate category scale(s) for that word is incremented. At the end of this procedure, a final score is assigned to each category, representing the percentage of words in the text sample matching that category. Importantly, the classification system includes categories tapping a variety of emotional dimensions, making it sensitive to differences among discrete emotions, including anxiety and anger (Kahn et al. 2007). Therefore, the software has seen increasing use as a measure of emotional disclosure (e.g., Bantum and Owen 2009; Pennebaker and Stone 2003).

In order to examine the emotional content of our 7,322 merchant reviews, each was submitted to LIWC analysis. The classification categories ‘anxiety’ and ‘anger’ represented the key variables of interest. Across all reviews in the set, the maximum value obtained for anxiety or anger was 50, and the average values for both categories were below 0.2 (this is unsurprising given use of a pre-defined dictionary which does not take context into consideration). 9.81% of the reviews contained at least one anxiety word, and 11.84% of the reviews contained at least one anger word. Notably, only 2.57% of the reviews contained both anxiety and anger words. Examples of anxiety-embedded and anger-embedded reviews are presented in Table 15.

Table 15: Examples of Emotional Reviews

#	Anxiety-Embedded Reviews	Anger-Embedded Reviews
1	I had some doubts about the item I purchased, never got an answer neither the store or the manufacturer	Lied about availability of product for two weeks, indicating that it had been shipped when, in fact, it was on back-order. Customer service? Don't bother!
2	Lost order per customer representative. No explanation. Now I am worried that they will "find" the order and will have to return since I am ordering from another vendor.	These people SUCK. They stalled the order for days trying to get me to buy extra shipping and other crap. Then they screwed up and didn't ship me one of the TV's I ordered. They SUCK.
3	The product was "backordered". It was ordered over a month ago as a gift, good price but never recieved the item. Said they would refund my credit card in 72 hours, and its been over a week and no refund. Getting a little worried. They are quick to reply to e-mails, but no refund. Seems to be a good company on yahoo, will update if the refund is made. (4th of July Holiday)	Extremely disappointed and offended. My Miele machine broke after 10 uses. When I called the store today, I was told that I was an idiot and that I was wasting 11 minutes of the salesperson's time with my idiocy. Then he hung up on me. I am contacting Miele headquarters to complain as well. I will never do business with this store again, and if you don't want to get ripped off and abused, you shouldn't either.

Following prior literature examining review helpfulness scores (Korfiatis et al. 2008; Mudambi and Schuff 2010), our analysis controlled for a series of relevant variables, including rating and rating squared, review length, review reading difficulty, and certain store characteristics. (1) Rating refers to the star rating of a review; the more stars a review received, the more positive the review is. Rating ranged from 1 star to 5 stars, and the average rating for the reviews in the set was 3.29. (2) A quadratic term of star rating was included to account for the non-linear relationship between rating and helpfulness (Mudambi and Schuff 2010). (3) Review length was operationalized as the number of words in a review; a longer review often provides more total information, and

may thus be considered more helpful. The analyzed reviews had on average 69.82 words.

(4) To control for review reading difficulty, we calculated the Coleman–Liau Index,³ an estimate of the U.S. grade level that a student would need to have achieved in order to read and understand the text (Coleman and Liao 1975). On average, the reviews in our data set were written at a 10th grade level. Lastly, we controlled for the effects of store characteristics, including a store’s average rating and the count of all its prior ratings. The former captures the overall reputation of a store, while the latter captures popularity. The operationalization of all variables is summarized in Table 16.

Table 16: Variables and Operationalizations

Variable Type	Variable Level	#	Variable	Operationalization	Notes
DV	Individual Review	1	Review Helpfulness	# helpful_votes / # total_votes	Range: [0, 1]
IV	Individual Review	2	Anxiety	(# anxiety-related words / # words in a review) * 100	Range: [0, 100] Coded by LIWC
		3	Anger	(# anger-related words / # words in a review) * 100	
Control	Individual Review	4	Rating	# of stars	Range: [1, 5]
		5	Length	# of words	
		6	Reading Difficulty	Coleman-Liau Index	
	Store	7	Reputation	average rating	Range: [1, 5]
		8	Popularity	# of ratings in total	

³ Alternative metrics for reading difficulty include the Flesch Reading Ease scale and Flesch-Kincaid Grade Level, among others (Dubay 2004). Although none of these metrics is perfect, they do correlate with the perceived difficulty of reading text samples. The results reported later do not change when one of the alternative metrics is used.

4.5.3 Data Analysis and Results

Analysis was performed following the approach of Mudambi and Schuff (2010), by using Tobit regression to analyze all reviews meeting the criteria described above ($N = 7,322$). We deemed this approach appropriate for the following reasons. First, the dependent variable was censored in nature: because it was constructed as a ratio, its value was bounded in range. Second, there exists a potential selection bias, because not every review reader casts a helpfulness vote. More importantly, the probability of a review being voted on might be correlated with explanatory variables such as review rating, review length, etc. Therefore, a sample containing only voted reviews might be non-random, and least-squares estimation of this sample would produce biased estimates (Greene and Zhang 2003).⁴

Table 17 contains the results of our empirical analysis. In the analysis, all independent variables were standardized to unify presentation of the results. The analysis indicates a good fit, with a highly significant likelihood ratio ($p < 0.001$) and pseudo R^2 value of 0.239 (Veall and Zimmermann 1996).

⁴ As a robustness check, we analyzed all the reviews ($N = 154,834$) by employing Heckman's (1979) two-step sample selection model. The first step is a Probit "selection" equation that examines determinants of whether a review gets voted on. In the second step, determinants of review helpfulness are estimated using only the voted reviews, conditional on the first step. In other words, the sample selection problem is treated as a form of omitted-variables bias, which is corrected by utilizing information obtained from the "selection" equation in the first step (Kuan et al. 2011). Results were consistent with those presented here.

Table 17: Tobit Analysis Results for Final Review Pool

Dependent variable: Review Helpfulness (N = 7,322)

Variable	Coefficient	Standard Error	t-value	Sig.
Constant	-6.352	0.063	-101.06	0.000
Rating	-2.196	0.091	-24.09	0.000
Rating ²	1.859	0.096	19.40	0.000
Length	0.207	0.014	15.28	0.000
Reading Difficulty	-0.152	0.020	-7.69	0.000
Average rating	-0.605	0.012	-51.86	0.000
# of ratings	-1.080	0.027	-40.11	0.000
Anxiety	0.034	0.013	2.55	0.011
Anger	-0.013	0.014	-0.91	0.360

Log likelihood = -32801.424

Likelihood Ratio = 14433.519 ($p = 0.000$, $df = 8$)McKelvey & Zavoina's Pseudo $R^2 = 0.239$

Tobit regression results involving the control variables were largely consistent with prior literature. Both linear ($\beta = -2.196$, $p < 0.001$) and squared ($\beta = 1.859$, $p < 0.001$) coefficients of review rating were significant and in the expected direction: reviews with lower ratings and/or higher extremity were considered more helpful. Additionally, a review was considered as more helpful to the extent that it was longer ($\beta = 0.207$, $p < 0.001$), and less difficult to understand ($\beta = -0.152$, $p < 0.001$). Coefficients for average rating ($\beta = -0.605$, $p < 0.001$) and count of ratings ($\beta = -1.080$, $p < 0.001$) were significant and negative; that is, controlling for all other variables, reviews of a well-liked or popular retailer were considered less valuable.

To explore Hypothesis 1, we compared the coefficients of anxiety and anger. First, to determine whether the emotion measures actually improve the model, we conducted a partial (or incremental) F-test of the null hypotheses that the coefficient of

anxiety of both anxiety and anger equals zero ($H_0: \beta_{\text{anxiety}} = \beta_{\text{anger}} = 0$). Results demonstrated that anxiety and anger were jointly significant ($F(2, 154826) = 3.66, p < 0.05$) and should therefore be included. Next, to examine our hypothesis that the effects of anger and anxiety differ, we used a Wald test to compare the equality of the two coefficients. This test revealed that the coefficient for anxiety was significantly higher than that for anger ($F(1, 154826) = 5.86, p < 0.05$). As predicted by H1, reviews containing content indicative of anxiety were more helpful than those containing content indicative of anger.

4.5.4 Discussion

Utilizing actual review data from Yahoo! Shopping websites, Study 3 supplemented the first two studies by providing empirical evidence for our primary hypothesis. In the context of real seller reviews, words related to anxiety and words related to anger exhibited differential impact on the perceived helpfulness of the overall review.

The use of empirical methods in this study necessitated certain limitations that may in turn present alternative interpretations. Most notably, given that our design used naturally occurring reviews, anxious reviewers may have in fact produced content that was objectively more helpful. This interpretation is compatible both with our theory, which assumes that the naïve theories of reviewers are generally accurate, and with prior research on the effects of anxiety and anger (Tiedens and Linton 2001). Even though some of our control variables were intended to account for information value, we cannot definitely conclude that it played no role in our findings. Importantly, however, this

concern is not applicable to the experiments of Study 1 and 2, which held constant the informative content of the reviews and still obtained the predicted differences in perceived helpfulness, along with evidence for the underlying process.

4.6 General Discussion

Together, the experiments in the first two studies and the empirical investigation in Study 3 provide converging evidence for our framework. Extending beyond valence-based approaches ('good' and 'bad'), these studies demonstrated the differential impact of two discrete emotions on review helpfulness. Reviews containing content indicative of anxiety were considered more helpful than those containing content indicative of anger, and their differential impact was explained by beliefs regarding the cognitive effort of reviewers.

4.6.1 Theoretical Implications

In contrast to the current, cognition-dominated literature on review helpfulness, we are among the first attempts to explore the effects of emotions above and beyond their non-emotional counterparts. Management scholars have begun to recognize the important role of affect in decision-making (Loewenstein and Lerner 2003), and research dealing with emotions has exploded, most notably in marketing (Bagozzi et al. 1999; Cohen et al. 2008) and organizational behavior (Ashkanasy et al. 2002; Brief and Weiss 2002). On the other hand, within the IS field, affective issues are often overlooked (see Sun and Zhang 2006a for a review). For example, among prominent conceptual frameworks for the effectiveness of information and communication technologies such as the Technology

Acceptance Model (Davis 1989), Media Richness Theory (Daft et al. 1987), and Task Technology Fit theory (Goodhue and Thompson 1995), none directly include emotional components. On the other hand, a small contingent of IS scholars have started to incorporate affective factors into established theories and frameworks (Deng and Poole 2010; Sun and Zhang 2006b; Venkatesh 2000). In keeping with those who have advocated the study of emotions in IS research (see de Guinea and Markus 2009), we contribute to this burgeoning area by exploring specific roles of emotion in e-commerce.

Prior empirical investigations of online reviews have tended to focus on ratings and observable reviewer characteristics, leaving the textual content of reviews relatively unexplored. Supplementing this work, we utilized content analysis and an experimental approach to demonstrate that emotions inferred from the text of seller reviews can predict the perceived helpfulness of the reviews. Consistent with the implications of others taking this approach (Cao et al. 2011; Pavlou and Dimoka 2006), our findings reveal that the rich information embedded in the text itself can be useful in explaining what constitutes a ‘helpful’ review.

Our results raise important issues concerning the application of ‘negativity bias’ to online WOM. Extending the logic of negativity bias to the context of emotional content, one would assume that reviews with negative emotions would be more helpful. However, this valence-based approach cannot account for the distinct effects of emotions similar in valence (Fontaine et al. 2007). Both anxiety and anger are negative, high-arousal emotions; nevertheless, due to their distinct motivations and behavioral implications, we expected and observed that they would influence perceptions of review helpfulness in distinct ways. Within psychology, there have been loud calls to move

‘beyond valence’ in examining the effect of emotions (Lerner and Keltner 2000), and within IS, a few scholars have explored the distinct roles of discrete emotions in technology acceptance (Venkatesh 2000) and online trust (Hwang and Kim 2007). In line with this movement, our studies provided initial evidence that emotional reviews - even those of the same valence and containing the same objective information - create distinct perceptions of review helpfulness.

The limited IS literature in emotions often makes the simplifying assumption that positive (negative) emotions lead to positive (negative) outcomes. Building on the categorization of discrete emotions, we adopt a more nuanced view to explain how specific embedded emotions impact perceptions of a review. Moreover, we introduce perceived cognitive effort as a mediator, and demonstrate its mediating role in our experimental study. Thus, our research provides novel support for theories of emotion that relate underlying cognitive appraisals to distinct thought processes (Lerner and Tiedens 2006; Tiedens and Linton 2001). Although we focused on anger and anxiety, the underlying logic could be applied to predict the effects of a wide variety of emotions. Furthermore, although we focused on perceptions of review helpfulness, the underlying mechanism we describe is applicable to any number of important consumer perceptions (reviewer expertise, trustworthiness of retailers, etc.).

4.6.2 Practical Implications

Although review authors undoubtedly have numerous motivations, one of these is often the desire to assist future customers via ‘helpful’ information regarding a seller, transaction, or product. Negative reviews have the potential to influence the attitude and

behaviors of future customers to a greater extent than positive reviews (Cao et al. 2011; Chevalier and Mayzlin 2006). However, it should not be assumed that a more emotional negative review will be perceived as more helpful; rather, this conclusion must be qualified according to the specific emotions involved. For instance, we observe that an anger-embedded review is perceived as less helpful than an anxiety-embedded review, even if the substantive content of the review is held constant. As a result, ‘ranting’ about a bad experience may be counterproductive for reviewers seeking to positively influence the choices of other customers. Instead, dissatisfied reviewers would be well advised to either avoid explicit expressions of anger or, alternatively, provide additional informative content to counteract its implications. At a broader level, review platforms themselves might utilize our findings in developing writing guidelines to encourage more useful seller reviews. For example, the admonition by Epinions.com “Do not use offensive language or content” is common among sites providing reviewer instructions; while intended to maintain decorum, this guideline is actually consistent with our implications regarding anger. More generally, review platforms cannot reasonably expect writers to amplify or suppress specific emotions; instead, they may ask reviewers to freely express emotions, but to think carefully about their tone as well as their content (e.g., by taking the perspective of a future reader).

Various empirical studies have explored the helpfulness of product reviews and provided implications for manufacturers and retailers (Chevalier and Mayzlin 2006; Forman et al. 2008; Mudambi and Schuff 2010). Supplementing these studies, our work focuses specifically on seller reviews, which are increasingly important in the branding and differentiation of online merchants. Generally speaking, merchants are aware of the

need to be vigilant and proactive in dealing with negative reviews; and many third-party sites provide mechanisms for doing so (e.g., BizRate allows vendors to post a public response immediately below a review). Assuming that a merchant seeks to identify (and respond to) negative WOM that is particularly influential on prospective customers, it may be presumed that angry reviews deserve particular attention. However, our findings suggest that this intuition is erroneous, as angry reviews appear to be discounted by prospective customers because of their embedded emotion. In contrast, reviews expressing anxiety may represent a more urgent concern.

4.6.3 Limitations and Future Research

Although our studies examined two particular emotions - anxiety and anger - that are prevalent in seller reviews, other emotions are also common (disappointment, happiness, surprise, etc.). Based on our results, it is worth considering how the presence of these emotions affects perceptions of helpfulness, and whether cognitive effort or other mediators best explain their effects. An appraisal-based approach offers many intriguing possibilities: for example, despite their opposing valence, both disappointment and happiness are characterized as high in certainty (Smith and Ellsworth 1985). It would be interesting to observe whether the presence of either emotion in a review generates similar effects on reader perceptions.

Our framework emphasizes the mediating role of perceived cognitive effort in explaining the effects of review-embedded emotions; however, we acknowledge that other mediating processes may also be relevant. In particular, although we ruled out alternative explanations based on valence, attribution, and empathy in Studies 1-2, we do

not deny that these factors play a role in determining review helpfulness, and future studies might consider their interaction with emotional content directly. Second, although Study 2 confirmed our hypotheses even after controlling for arousal, our results suggest that arousal may also play a role in distinguishing the effects of various reviewer emotions. This possibility is certainly worthy of further exploration. Finally, anxiety and anger differ not only in underlying appraisals of certainty, but also in appraisals of control. Our model assumes that for readers assessing the usefulness of a review, certainty is the more directly applicable appraisal. However, for other aspects of assessment, control appraisals may be especially useful in explaining differential effects of anxiety and anger.

Third, the studies in this paper were conducted exclusively in the context of seller reviews. Although there is a reason to believe that our underlying arguments will apply similarly to product reviews, additional factors may need to be considered. For example, in the case of a product review, the specific target of a reviewer's emotion may be unclear (the product itself, the manufacturer, retailer, etc.), limiting the ability of readers to draw inferences. Moreover, seller reviewers are generally anonymous, whereas product reviewers are often identifiable in terms of expertise, purchase history, demographics, etc. The availability of this relevant information may weaken the extent to which emotional cues affect inferences about the reviewer. Therefore, future research is needed to extend our investigation to a product review setting.

Finally, two assumptions of our framework merit further examination. First, we stress the impact of reviewers' emotional state on their cognitive effort; however, in keeping with longstanding debate regarding the relative precedence of affect and

cognition, it is also plausible that cognitive cues (such as effort) influence the emotional state of the reviewer. In our studies, we assume that the former route is more applicable, because the indicated emotion has already resulted from interaction with the seller. However, our main argument - that readers connect anger/anxiety with less/more cognitive effort - does not depend on directionality. Second, despite evidence that individuals attend to and recognize emotional cues in verbal communication (Lindquist et al. 2006; Zeelenberg et al. 2006), we acknowledge that readers will not always be accurate in identifying the emotional state of the author. Although this error works against our hypotheses (making our studies more conservative), it is an important caveat; indeed, the ability of readers to accurately recognize reviewer emotions is a topic worth further exploration.

4.7 Conclusion

In keeping with recent interest in incorporating affective factors into existing IS frameworks, we believe that scholars will benefit strongly from a better understanding of the impact of discrete emotions. Our research provide both experimental and empirical evidence that negative seller reviews containing diverse emotions are not ‘created equal,’ but rather have differential effects on the perceived usefulness of peer information. As such, we believe that this work extends current understanding of an under-studied phenomenon, and we look forward to further research exploring causes and consequences of emotions in the online environment.

APPENDIX A

FOR CHAPTER 2: DEPENDENT VARIABLES AND COVARIATES

Trust in vendor: (Gefen 2000)

- I believe that Store X is trustworthy.
- Even if Store X was not monitored, I would trust them to do the job right.**
- I trust Store X.

Intention to purchase: (Jarvenpaa et al. 2000; Pavlou 2003)

- For this purchase, how likely is it that you will buy from Store X?
- If you need to buy an electronic product in the future, how likely is it that you would consider making another purchase from Store X?**
- How likely is it that you would transact with Store X in the near future?

Propensity to Trust: (Gefen 2000)

- I generally trust other people.
- I tend to count upon other people.
- I generally have faith in humanity.
- I feel that people are generally reliable.
- I generally trust other people unless they give me reasons not to.

Institution-based Trust: (McKnight et al. 2002)

- I feel good about how things go when I do purchasing or other activities on the Internet.
- I am comfortable making purchases on the Internet.

Notes:

* “Store X” was replaced with the corresponding store’s name in experiments.

** The item was not used in Study 3.

APPENDIX B

FOR CHAPTER 3: VARIABLES MEASURED IN THE EXPERIMENTS

Helpfulness: (Sen and Lerman 2007)

Using the scales below, how would you describe the above user review?

- not at all helpful / very helpful
- not at all useful / very useful
- not at all informative / very informative

Perceived rationality of reviewers: (Tormala et al. 2011)

- In your opinion, how rational was the reviewer in writing this review about the app?

<not rational at all / very rational>

- In your opinion, to what extent was this review based on logic and reason? <not at all / very much>

Valence: (Mackenzie and Lutz 1989)

Overall, how would you describe the user's feelings regarding the experience he/she wrote in the review above?

- very bad / very good
- very unfavorable / very favorable
- very unpleasant / very pleasant

Arousal: (Berger 2011)

Using the scales below, how do you think the reviewer was feeling at the time that he/she wrote the review?

- very passive / very active
- very mellow / very fired up
- very low energy level / very high energy level

APPENDIX C

FOR CHAPTER 4: VARIABLES MEASURED IN THE EXPERIMENTS

Helpfulness: (Sen and Lerman 2007)

Using the scales below, how would you describe the above consumer review?

- not at all helpful / very helpful
- not at all useful / very useful
- not at all informative / very informative

Perceived cognitive effort of reviewers: (Huddy et al. 2007)

- In your opinion, how much effort had the reviewer put into writing this review?
- In your opinion, how much thought had the reviewer given to the above review when he/she wrote it?
- In your opinion, how much time did the reviewer spent writing this review?

Valence: (Mackenzie and Lutz 1989)

Overall, how would you describe the customer's feelings regarding the experience he/she wrote in the review above?

- very bad / very good
- very unfavorable / very favorable
- very unpleasant / very pleasant

Arousal: (Berger 2011)

Using the scales below, how do you think the reviewer was feeling at the time that he/she wrote the review?

- very passive / very active
- very mellow / very fired up
- very low energy level / very high energy level

Attribution about the reviewer: (Sen and Lerman 2007)

There are a wide variety of reasons that customers might write a store review. Rate the extent to which you agree with the following statements.

- The cause of the review was something about the reviewer.

Empathy: (McCullough et al. 1997; Toi and Batson 1982)

- While reading this review, to what extent did you feel like you were experiencing the same emotions as the reviewer?
- While reading this review, to what extent did you feel concerned for the reviewer?
- While reading this review, to what extent did you feel moved by the review?

REFERENCES

- Ahluwalia, R. 2002. "How Prevalent Is the Negativity Effect in Consumer Environments?," *Journal of Consumer Research* (29:2), Sep, pp 270-279.
- Apple. 2008. "Apple Introduces the New Iphone 3g." Retrieved April 20, 2012, from <http://www.apple.com/pr/library/2008/06/09Apple-Introduces-the-New-iPhone-3G.html>
- Apple. 2012. "Adhering to Guidelines on Third-Party Marketing Services." Retrieved April 20, 2012, from <https://developer.apple.com/news/index.php?id=02062012a>
- Ashkanasy, N.M., Hartel, C.E.J., and Daus, C.S. 2002. "Diversity and Emotion: The New Frontiers in Organizational Behavior Research," *Journal of Management* (28:3), pp 307-338.
- Atkinson, A.P., Dittrich, W.H., Gemmell, A.J., and Young, A.W. 2004. "Emotion Perception from Dynamic and Static Body Expressions in Point-Light and Full-Light Displays," *Perception* (33:6), pp 717-746.
- Ba, S., and Pavlou, P.A. 2002. "Evidence of the Effect of Trust Building Technology in Electronic Markets: Price Premiums and Buyer Behavior," *MIS Quarterly* (26:3), Sep, pp 243-268.
- Bagozzi, R.P., Gopinath, M., and Nyer, P.U. 1999. "The Role of Emotions in Marketing," *Journal of the Academy of Marketing Science* (27:2), pp 184-206.
- Bagozzi, R.P., and Phillips, L.W. 1982. "Representing and Testing Organizational Theories - a Holistic Construal," *Administrative Science Quarterly* (27:3), pp 459-489.
- Bantum, E.O., and Owen, J.E. 2009. "Evaluating the Validity of Computerized Content Analysis Programs for Identification of Emotional Expression in Cancer Narratives," *Psychological Assessment* (21:1), Mar, pp 79-88.
- Baron, R.M., and Kenny, D.A. 1986. "The Moderator Mediator Variable Distinction in Social Psychological-Research - Conceptual, Strategic, and Statistical

- Considerations," *Journal of Personality and Social Psychology* (51:6), Dec, pp 1173-1182.
- Barrett, L.F., and Russell, J.A. 1998. "Independence and Bipolarity in the Structure of Current Affect," *Journal of Personality and Social Psychology* (74:4), Apr, pp 967-984.
- Baumeister, R.F., Bratslavsky, E., Finkenauer, C., and Vohs, K.D. 2001. "Bad Is Stronger Than Good," *Review of General Psychology* (5:4), pp 323-370.
- Berger, J. 2011. "Arousal Increases Social Transmission of Information," *Psychological Science* (22:7), Jul, pp 891-893.
- Berger, J., and Milkman, K.L. 2012. "What Makes Online Content Viral?," *Journal of Marketing Research* (49:2), Apr, pp 192-205.
- Berkowitz, L., and Harmon-Jones, E. 2004. "Toward an Understanding of the Determinants of Anger," *Emotion* (4:2), Jun, pp 107-130.
- Bickart, B., and Schindler, R.M. 2001. "Internet Forums as Influential Sources of Consumer Information," *Journal of Interactive Marketing* (15:3), pp 31-40.
- Birnbaum, M.H. 1974. "Nonadditivity of Personality Impressions," *Journal of Experimental Psychology* (102:3), pp 543-561.
- Bodenhausen, G.V., Sheppard, L.A., and Kramer, G.P. 1994. "Negative Affect and Social Judgment - the Differential Impact of Anger and Sadness," *European Journal of Social Psychology* (24:1), Jan-Feb, pp 45-62.
- Bond, S.D., Bettman, J.R., and Luce, M.F. 2008. "Consumer Judgment from a Dual-Systems Perspective: Recent Evidence and Emerging Issues," in: *Review of Marketing Research*, N.K. Malhotra (ed.). Armonk, NY: M. E. Sharpe.
- Brief, A.P., and Weiss, H.M. 2002. "Organizational Behavior: Affect in the Workplace," *Annual Review of Psychology* (53), pp 279-307.
- Brosch, T., Pourtois, G., and Sander, D. 2010. "The Perception and Categorisation of Emotional Stimuli: A Review," *Cognition & Emotion* (24:3), pp 377-400.

- Buck, R. 1988. *Human Motivation and Emotion*, (2 ed.). Oxford, England: John Wiley & Sons.
- Burgoon, J.K. 1993. "Interpersonal Expectations, Expectancy Violations, and Emotional Communication," *Journal of Language and Social Psychology* (12:1-2), pp 30-48.
- Cao, Q., Duan, W.J., and Gan, Q.W. 2011. "Exploring Determinants of Voting for the "Helpfulness" of Online User Reviews: A Text Mining Approach," *Decision Support Systems* (50:2), Jan, pp 511-521.
- Chaiken, S., and Trope, Y. 1999. *Dual-Process Theories in Social Psychology*. The Guilford Press.
- Chen, P.-Y., Dhanasobhon, S., and Smith, M.D. 2008. "All Reviews Are Not Created Equal: The Disaggregate Impact of Reviews and Reviewers at Amazon.Com," in: *SSRN eLibrary*.
- Chen, Y.B., Wang, Q., and Xie, J.H. 2011. "Online Social Interactions: A Natural Experiment on Word of Mouth Versus Observational Learning," *Journal of Marketing Research* (48:2), Apr, pp 238-254.
- Chevalier, J.A., and Mayzlin, D. 2006. "The Effect of Word of Mouth on Sales: Online Book Reviews," *Journal of Marketing Research* (43:3), Aug, pp 345-354.
- Cohen, J.B., Pham, M.T., and Andrade, E.B. 2008. "The Nature and Role of Affect in Consumer Behavior," in: *Handbook of Consumer Psychology*, C.P. Haugtvedt, P. Herr and F. Kardes (eds.). Mahwah, NJ: Lawrence Erlbaum, pp. 297-348.
- Coleman, M., and Liao, T.L. 1975. "A Computer Readability Formula Designed for Machine Scoring," *Journal of Applied Psychology* (60:2), pp 283-284.
- Connors, L., Mudambi, S.M., and Schuff, D. 2011. "Is It the Review or the Reviewer? A Multi-Method Approach to Determine the Antecedents of Online Review Helpfulness," in: *Proceedings of the 2011 44th Hawaii International Conference on System Sciences*. IEEE Computer Society, pp. 1-10.
- Dabholkar, P.A. 2006. "Factors Influencing Consumer Choice of a "Rating Web Site": An Experimental Investigation of an Online Interactive Decision Aid," *Journal of Marketing Theory & Practice* (14:4), Fall2006, pp 259-273.

- Daft, R.L., Lengel, R.H., and Trevino, L.K. 1987. "Message Equivocality, Media Selection, and Manager Performance - Implications for Information-Systems," *MIS Quarterly* (11:3), Sep, pp 355-366.
- Danesh, H.B. 1977. "Anger and Fear," *American Journal of Psychiatry* (134:10), pp 1109-1112.
- Davies, J., and Brember, I. 1994. "The Reliability and Validity of the Esmileyíscale," *British Educational Research Journal* (20:4), pp 447-454.
- Davis, F.D. 1989. "Perceived Usefulness, Perceived Ease of Use, and User Acceptance of Information Technology," *MIS Quarterly* (13:3), Sep, pp 319-340.
- de Guinea, A.O., and Markus, M.L. 2009. "Why Break the Habit of a Lifetime? Rethinking the Roles of Intention, Habit, and Emotion in Continuing Information Technology Use," *MIS Quarterly* (33:3), Sep, pp 433-444.
- Dellarocas, C. 2003. "The Digitization of Word of Mouth: Promise and Challenges of Online Feedback Mechanisms," *Management Science* (49:10), Oct, pp 1407-1424.
- Deng, L.Q., and Poole, M.S. 2010. "Affect in Web Interfaces: A Study of the Impacts of Web Page Visual Complexity and Order," *MIS Quarterly* (34:4), Dec, pp 711-730.
- Dijksterhuis, A., and Aarts, H. 2003. "On Wildebeests and Humans: The Preferential Detection of Negative Stimuli," *Psychological Science* (14:1), Jan, pp 14-18.
- DuBay, W.H. 2004. "The Principles of Readability." *Impact Information*, from <http://www.nald.ca/library/research/readab/readab.pdf>
- Ekman, P. 1992. "Are There Basic Emotions," *Psychological Review* (99:3), Jul, pp 550-553.
- Ekman, P., and Friesen, W.V. 1971. "Constants across Cultures in Face and Emotion," *Journal of Personality and Social Psychology* (17:2), pp 124-&.

- Ellsworth, P.C., and Scherer, K.R. 2003. "Appraisal Processes in Emotion," in: *Handbook of Affective Sciences.*, R.J. Davidson, K.R. Scherer and H.H. Goldsmith (eds.). New York, NY, US: Oxford University Press, pp. 572-595.
- Eysenck, M.W. 1982. *Attention and Arousal, Cognition and Performance*. New York: Springer.
- Fantino, E., Kasdon, D., and Stringer, N. 1970. "The Yerkes-Dodson Law and Alimentary Motivation," *Canadian Journal of Psychology/Revue canadienne de psychologie* (24:2), pp 77-84.
- Fedorikhin, A., and Patrick, V.M. 2010. "Positive Mood and Resistance to Temptation: The Interfering Influence of Elevated Arousal," *Journal of Consumer Research* (37:4), Dec, pp 698-711.
- Feldman, J.M. 1999. "Four Questions About Human Social Behavior," in: *Social Psychology and Cultural Context*, J. Adamopoulos and Y. Kashima (eds.). CA: Sage Publications, pp. 43-62.
- Feldman, J.M., and Lynch, J.G. 1988. "Self-Generated Validity and Other Effects of Measurement on Belief, Attitude, Intention, and Behavior," *Journal of Applied Psychology* (73:3), Aug, pp 421-435.
- Ferrin, D.L., and Dirks, K.T. 2003. "The Use of Rewards to Increase and Decrease Trust: Mediating Processes and Differential Effects," *Organization Science* (14:1), Jan-Feb, pp 18-31.
- Ferrin, D.L., Kim, P.H., Cooper, C.D., and Dirks, K.T. 2007. "Silence Speaks Volumes: The Effectiveness of Reticence in Comparison to Apology and Denial for Responding to Integrity- and Competence-Based Trust Violations," *Journal of Applied Psychology* (92:4), Jul, pp 893-908.
- Fishbein, M., and Ajzen, I. 1975. *Belief, Attitude, Intention, and Behavior: An Introduction to Theory and Research*. Reading, MA: Addison-Wesley Publishing Company.
- Fiske, S.T. 1980. "Attention and Weight in Person Perception - the Impact of Negative and Extreme Behavior," *Journal of Personality and Social Psychology* (38:6), pp 889-906.

- Fontaine, J.R.J., Scherer, K.R., Roesch, E.B., and Ellsworth, P.C. 2007. "The World of Emotions Is Not Two-Dimensional," *Psychological Science* (18:12), Dec, pp 1050-1057.
- Forgas, J.P. 1995. "Mood and Judgment - the Affect Infusion Model (Aim)," *Psychological Bulletin* (117:1), Jan, pp 39-66.
- Forman, C., Ghose, A., and Wiesenfeld, B. 2008. "Examining the Relationship between Reviews and Sales: The Role of Reviewer Identity Disclosure in Electronic Markets," *Information Systems Research* (19:3), pp 291-313.
- Frijda, N.H. 1986. *The Emotions*. Cambridge, England: Cambridge University Press.
- Frijda, N.H. 1993. "Moods, Emotion Episodes, and Emotions," in: *Handbook of Emotions*. New York, NY, US: Guilford Press, pp. 381-403.
- Frijda, N.H. 2000. "Emotions," in: *International Handbook of Psychology*, K. Pawlik and M. Rosenzweig (eds.). London: Sage, pp. 207-222.
- Fulcher, E. 2003. *Cognitive Psychology*. Learning Matters.
- Gefen, D. 2000. "E-Commerce: The Role of Familiarity and Trust," *Omega-International Journal of Management Science* (28:6), Dec, pp 725-737.
- Goodhue, D.L., and Thompson, R.L. 1995. "Task-Technology Fit and Individual-Performance," *MIS Quarterly* (19:2), Jun, pp 213-236.
- Gorn, G., Pham, M.T., and Sin, L.Y. 2001. "When Arousal Influences Ad Evaluation and Valence Does Not (and Vice Versa)," *Journal of Consumer Psychology* (11:1), pp 43-55.
- Greene, W.H., and Zhang, C. 2003. *Econometric Analysis*. Prentice hall Upper Saddle River, NJ.
- Han, S., Lerner, J.S., and Keltner, D. 2007. "Feelings and Consumer Decision Making: The Appraisal-Tendency Framework," *Journal of Consumer Psychology* (17:3), pp 158-168.

- Hanoch, Y., and Vitouch, O. 2004. "When Less Is More - Information, Emotional Arousal and the Ecological Reframing of the Yerkes-Dodson Law," *Theory & Psychology* (14:4), Aug, pp 427-452.
- Heckman, J.J. 1979. "Sample Selection Bias as a Specification Error," *Econometrica* (47:1), pp 153-161.
- Heider, F. 1958. *The Psychology of Interpersonal Relations*. Lawrence Erlbaum Associates.
- Heilman, K.M. 1997. "The Neurobiology of Emotional Experience," *Journal of Neuropsychiatry and Clinical Neurosciences* (9:3), Sum, pp 439-448.
- Heneman, H.G., and Judge, T. 2006. *Staffing Organizations*, (5 ed.). Boston: McGraw-Hill Irwin.
- Hennig-Thurau, T., Gwinner, K.P., Walsh, G., and Gremler, D.D. 2004. "Electronic Word-of-Mouth Via Consumer-Opinion Platforms: What Motivates Consumers to Articulate Themselves on the Internet?," *Journal of Interactive Marketing* (18:1), pp 38-52.
- Huddy, L., Feldman, S., and Cassese, E. 2007. "On the Distinct Political Effects of Anxiety and Anger," in: *The Affect Effect: Dynamics of Emotion in Political Thinking and Behavior*, A. Crigler, M. MacKuen, G.E. Marcus and W.R. Neuman (eds.). University of Chicago Press, pp. 202-230.
- Humphreys, M.S., and Revelle, W. 1984. "Personality, Motivation, and Performance - a Theory of the Relationship between Individual-Differences and Information-Processing," *Psychological Review* (91:2), pp 153-184.
- Hwang, Y.J., and Kim, D.J. 2007. "Customer Self-Service Systems: The Effects of Perceived Web Quality with Service Contents on Enjoyment, Anxiety, and E-Trust," *Decision Support Systems* (43:3), Apr, pp 746-760.
- Jarvenpaa, S.L., Tractinsky, N., and Vitale, M. 2000. "Consumer Trust in an Internet Store," *Information Technology and Management* (1:1-2), pp 45-71.
- Johnston, W.A., and Heinz, S.P. 1978. "Flexibility and Capacity Demands of Attention," *Journal of Experimental Psychology-General* (107:4), pp 420-435.

- Jones, E.E., and Davis, K.E. 1965. "From Acts to Dispositions - the Attribution Process in Person Perception," *Advances in experimental social psychology* (2:4), pp 219-266.
- Jones, E.E., and Davis, K.E. 1966. "From Acts to Dispositions the Attribution Process in Person Perception," *Advances in experimental social psychology* (2), pp 219-266.
- Jones, Q., Ravid, G., and Rafaeli, S. 2004. "Information Overload and the Message Dynamics of Online Interaction Spaces: A Theoretical Model and Empirical Exploration," *Information Systems Research* (15:2), Jun, pp 194-210.
- Judd, C.M., Kenny, D.A., and McClelland, G.H. 2001. "Estimating and Testing Mediation and Moderation in within-Subject Designs," *Psychological Methods* (6:2), Jun, pp 115-134.
- Kahn, J.H., Tobin, R.M., Massey, A.E., and Anderson, J.A. 2007. "Measuring Emotional Expression with the Linguistic Inquiry and Word Count," *American Journal of Psychology* (120:2), Sum, pp 263-286.
- Kahneman, D. 1973. *Attention and Effort*. Englewood Cliffs, NJ: Prentice-Hall.
- Kahneman, D., and Tversky, A. 1979. "Prospect Theory - Analysis of Decision under Risk," *Econometrica* (47:2), pp 263-291.
- Kaneshige, T. 2010. "Apple App Store's Dirty Little Secret." *PCWorld* Retrieved April 12, 2012, from http://www.pcworld.com/article/189973/apple_app_stores_dirty_little_secret.html
- Kaufman, B.E. 1999. "Emotional Arousal as a Source of Bounded Rationality," *Journal of Economic Behavior & Organization* (38:2), Feb, pp 135-144.
- Kelley, H.H. 1967. "Attribution Theory in Social Psychology," University of Nebraska Press.
- Keltner, D., Ellsworth, P.C., and Edwards, K. 1993. "Beyond Simple Pessimism - Effects of Sadness and Anger on Social-Perception," *Journal of Personality and Social Psychology* (64:5), May, pp 740-752.

- Keltner, D., and Haidt, J. 1999. "Social Functions of Emotions at Four Levels of Analysis," *Cognition & Emotion* (13:5), Sep, pp 505-521.
- Khare, A., Labrecque, L.I., and Asare, A.K. 2011. "The Assimilative and Contrastive Effects of Word-of-Mouth Volume: An Experimental Examination of Online Consumer Ratings," *Journal of Retailing* (87:1), Mar, pp 111-126.
- Kim, P.H., Dirks, K.T., Cooper, C.D., and Ferrin, D.L. 2006. "When More Blame Is Better Than Less: The Implications of Internal Vs. External Attributions for the Repair of Trust after a Competence- Vs. Integrity-Based Trust Violation," *Organizational Behavior and Human Decision Processes* (99:1), Jan, pp 49-65.
- Kim, P.H., Ferrin, D.L., Cooper, C.D., and Dirks, K.T. 2004. "Removing the Shadow of Suspicion: The Effects of Apology Versus Denial for Repairing Competence- Versus Integrity-Based Trust Violations," *Journal of Applied Psychology* (89:1), Feb, pp 104-118.
- Korfiatis, N., Rodriguez, D., and Sicilia, M.A. 2008. "The Impact of Readability on the Usefulness of Online Product Reviews: A Case Study on an Online Bookstore," in: *Emerging Technologies and Information Systems for the Knowledge Society, Proceedings*, M.D. Lytras, J.M. Carroll, E. Damiani and R.D. Tennyson (eds.). Berlin: Springer-Verlag Berlin, pp. 423-432.
- Kotler, P., and Armstrong, G. 2010. *Principles of Marketing*. Prentice Hall.
- Kousta, S.T., Vinson, D.P., and Vigliocco, G. 2009. "Emotion Words, Regardless of Polarity, Have a Processing Advantage over Neutral Words," *Cognition* (112:3), Sep, pp 473-481.
- Krueger, J.I., and Funder, D.C. 2004. "Towards a Balanced Social Psychology: Causes, Consequences, and Cures for the Problem-Seeking Approach to Social Behavior and Cognition," *Behavioral and Brain Sciences* (27:3), Jun, pp 313-+.
- Kuan, K.K.Y., Hui, K.-L., Prasarnphanich, P., and Lai, J.H.Y. 2011. "Online Voting and Review Helpfulness: An Empirical Investigation." Working paper.
- Larsen, R.J., and Diener, E. 1992. "Promises and Problems with the Circumplex Model of Emotion," *Review of Personality and Social Psychology* (13:13), pp 25-59.

- Lazarus, R.S. 1982. "Thoughts on the Relations between Emotion and Cognition," *American Psychologist* (37:9), 1982, pp 1019-1024.
- Lazarus, R.S. 1991. *Emotion and Adaptation*. Oxford University Press, USA.
- Lerner, J.S., Goldberg, J.H., and Tetlock, P.E. 1998. "Sober Second Thought: The Effects of Accountability, Anger, and Authoritarianism on Attributions of Responsibility," *Personality and Social Psychology Bulletin* (24:6), Jun, pp 563-574.
- Lerner, J.S., Gonzalez, R.M., Small, D.A., and Fischhoff, B. 2003. "Effects of Fear and Anger on Perceived Risks of Terrorism: A National Field Experiment," *Psychological Science* (14:2), Mar, pp 144-150.
- Lerner, J.S., and Keltner, D. 2000. "Beyond Valence: Toward a Model of Emotion-Specific Influences on Judgement and Choice," *Cognition & Emotion* (14:4), Jul, pp 473-493.
- Lerner, J.S., and Keltner, D. 2001. "Fear, Anger, and Risk," *Journal of Personality and Social Psychology* (81:1), Jul, pp 146-159.
- Lerner, J.S., and Tiedens, L.Z. 2006. "Portrait of the Angry Decision Maker: How Appraisal Tendencies Shape Anger's Influence on Cognition," *Journal of Behavioral Decision Making* (19:2), Apr, pp 115-137.
- Levenson, R.W. 1992. "Autonomic Nervous-System Differences among Emotions," *Psychological Science* (3:1), Jan, pp 23-27.
- Lindquist, K.A., Barrett, L.F., Bliss-Moreau, E., and Russell, J.A. 2006. "Language and the Perception of Emotion," *Emotion* (6:1), Feb, pp 125-138.
- Liu, Y., Huang, X.J., Ani, A.J., and Yu, X.H. 2008. "Modeling and Predicting the Helpfulness of Online Reviews," in: *Icdm 2008: Eighth Ieee International Conference on Data Mining, Proceedings*, D. Gunopulos, F. Turini, C. Zaniolo, N. Ramakrishnan and X.D. Wu (eds.). Los Alamitos: Ieee Computer Soc, pp. 443-452.

- Loewenstein, G., and Lerner, J.S. 2003. "The Role of Affect in Decision Making," in: *Handbook of Affective Sciences.*, R.J. Davidson, K.R. Scherer and H.H. Goldsmith (eds.). New York, NY, US: Oxford University Press, pp. 619-642.
- Lord, R.G., and Kanfer, R. 2002. "Emotions and Organizational Behavior," in: *Emotions in the Workplace*, R.G. Lord, R.J. Klimoski and R. Kanfer (eds.). San Francisco, CA: Jossey-Bass, pp. 5-19.
- Mackenzie, S.B., and Lutz, R.J. 1989. "An Empirical-Examination of the Structural Antecedents of Attitude toward and Ad in an Advertising Pretesting Context," *Journal of Marketing* (53:2), Apr, pp 48-65.
- Mano, H. 1991. "The Structure and Intensity of Emotional Experiences - Method and Context Convergence," *Multivariate Behavioral Research* (26:3), pp 389-411.
- Manstead, A.S.R., and Fischer, A.H. 2001. "Social Appraisal: The Social World as Object of and Influence on Appraisal Processes," in: *Appraisal Processes in Emotion: Theory, Research, Application*, K.R. Scherer, A. Schorr and T. Johnstone (eds.). New York: Oxford University Press, pp. 221-232.
- Martijn, C., Spears, R., Vanderpligt, J., and Jakobs, E. 1992. "Negativity and Positivity Effects in Person Perception and Inference - Ability Versus Morality," *European Journal of Social Psychology* (22:5), Sep-Oct, pp 453-463.
- Mayer, R.C., Davis, J.H., and Schoorman, F.D. 1995. "An Integrative Model of Organizational Trust," *Academy of Management Review* (20:3), Jul, pp 709-734.
- McCullough, M.E., Worthington, E.L., and Rachal, K.C. 1997. "Interpersonal Forgiving in Close Relationships," *Journal of Personality and Social Psychology* (73:2), Aug, pp 321-336.
- McKnight, D.H., Choudhury, V., and Kacmar, C. 2002. "Developing and Validating Trust Measures for E-Commerce: An Integrative Typology," *Information Systems Research* (13:3), Sep, pp 334-359.
- McKnight, D.H., Cummings, L.L., and Chervany, N.L. 1998. "Initial Trust Formation in New Organizational Relationships," *Academy of Management Review* (23:3), Jul, pp 473-490.

- Mudambi, S.M., and Schuff, D. 2010. "What Makes a Helpful Online Review? A Study of Customer Reviews on Amazon.Com," *MIS Quarterly* (34:1), Mar, pp 185-200.
- Mueller, J.H. 1979. "Anxiety and Encoding Processes in Memory," *Personality and Social Psychology Bulletin* (5:3), pp 288-294.
- Murphy, G.L., and Medin, D.L. 1985. "The Role of Theories in Conceptual Coherence," *Psychological Review* (92:3), pp 289-316.
- Niedenthal, P.M. 2008. "Emotion Concepts," in: *Handbook of Emotions*, M. Lewis, J.M. Haviland-Jones and L.F. Barrett (eds.). New York: The Guilford Press, pp. 587-600.
- Nunnally, J. 1967. *Psychometric Methods*. New York: McGraw-Hill.
- O'Rourke, D. 2004. "The Measurement of Pain in Infants, Children, and Adolescents: From Policy to Practice," *Physical Therapy* (84:6), Jun, pp 560-570.
- Oatley, K., and Johnson-laird, P.N. 1987. "Towards a Cognitive Theory of Emotions," *Cognition & Emotion* (1:1), pp 29-50.
- Pavlou, P.A. 2003. "Consumer Acceptance of Electronic Commerce: Integrating Trust and Risk with the Technology Acceptance Model," *International Journal of Electronic Commerce* (7:3), Spr, pp 101-134.
- Pavlou, P.A., and Dimoka, A. 2006. "The Nature and Role of Feedback Text Comments in Online Marketplaces: Implications for Trust Building, Price Premiums, and Seller Differentiation," *Information Systems Research* (17:4), Dec, pp 392-414.
- Pennebaker, J.W., Booth, R.J., and Francis, M.E. 2007. *Linguistic Inquiry and Word Count (Liw2007)*. Austin, TX: LIWC (<http://www.liwc.net>).
- Pennebaker, J.W., and Francis, M.E. 1996. "Cognitive, Emotional, and Language Processes in Disclosure," *Cognition & Emotion* (10:6), Nov, pp 601-626.
- Pennebaker, J.W., and Stone, L.D. 2003. "Words of Wisdom: Language Use over the Life Span," *Journal of Personality and Social Psychology* (85:2), Aug, pp 291-301.

- Qu, Z., Zhang, H., and Li, H.Z. 2008. "Determinants of Online Merchant Rating: Content Analysis of Consumer Comments About Yahoo Merchants," *Decision Support Systems* (46:1), Dec, pp 440-449.
- Raghunathan, R., and Pham, M.T. 1999. "All Negative Moods Are Not Equal: Motivational Influences of Anxiety and Sadness on Decision Making," *Organizational Behavior and Human Decision Processes* (79), pp 56-77.
- Resnick, P., and Zeckhauser, R. 2002. "Trust among Strangers in Internet Transactions: Empirical Analysis of Ebay's Reputation System," in: *The Economics of the Internet and E-Commerce*, M.R. Baye (ed.). Greenwich, CT: JAI Press, pp. 127-157.
- Roseman, I.J. 1984. "Cognitive Determinants of Emotion: A Structural Theory," *Review of Personality and Social Psychology* (5), pp 11-36.
- Roseman, I.J., and Smith, C.A. 2001. "Appraisal Theory: Overview, Assumptions, Varieties, Controversies," in: *Appraisal Processes in Emotion: Theory, Methods, Research*, K.R. Scherer, A. Schorr and T. Johnstone (eds.). New York, NY, US: Oxford University Press, pp. 3-19.
- Rothenberg, A. 1971. "On Anger," *American Journal of Psychiatry* (128:4), pp 454-460.
- Rozin, P., and Royzman, E.B. 2001. "Negativity Bias, Negativity Dominance, and Contagion," *Personality and Social Psychology Review* (5:4), pp 296-320.
- Russell, J.A. 1980. "A Circumplex Model of Affect," *Journal of Personality and Social Psychology* (39:6), pp 1161-1178.
- Russell, J.A., and Barrett, L.F. 1999. "Core Affect, Prototypical Emotional Episodes, and Other Things Called Emotion: Dissecting the Elephant," *Journal of Personality and Social Psychology* (76:5), May, pp 805-819.
- Russell, J.A., Lewicka, M., and Niit, T. 1989. "A Cross-Cultural-Study of a Circumplex Model of Affect," *Journal of Personality and Social Psychology* (57:5), Nov, pp 848-856.
- Russell, J.A., and Mehrabian, A. 1977. "Evidence for a Three-Factor Theory of Emotions," *Journal of Research in Personality* (11:3), pp 273-294.

- Sanbonmatsu, D.M., and Kardes, F.R. 1988. "The Effects of Physiological Arousal on Information-Processing and Persuasion," *Journal of Consumer Research* (15:3), Dec, pp 379-385.
- Scherer, K.R. 2003. "Introduction: Cognitive Components of Emotion," in: *Handbook of the Affective Sciences*, R.J. Davidson, K.R. Scherer and H.H. Goldsmith (eds.). New York and Oxford: Oxford University Press, pp. 563-571.
- Scherer, K.R., and Grandjean, D. 2008. "Facial Expressions Allow Inference of Both Emotions and Their Components," *Cognition & Emotion* (22:5), pp 789-801.
- Scherer, K.R., Schorr, A., and Johnstone, T. 2001. *Appraisal Processes in Emotion: Theory, Methods, Research*. New York, NY, US: Oxford University Press.
- Sen, S., and Lerman, D. 2007. "Why Are You Telling Me This? An Examination into Negative Consumer Reviews on the Web," *Journal of Interactive Marketing* (21:4), Fal, pp 76-94.
- Shapiro, S., MacInnis, D.J., and Park, C.W. 2002. "Understanding Program-Induced Mood Effects: Decoupling Arousal from Valence," *Journal of Advertising* (31:4), Win, pp 15-26.
- Shaver, P., Schwartz, J., Kirson, D., and Oconnor, C. 1987. "Emotion Knowledge - Further Exploration of a Prototype Approach," *Journal of Personality and Social Psychology* (52:6), Jun, pp 1061-1086.
- Siemer, M. 2008. "Beyond Prototypes and Classical Definitions: Evidence for a Theory-Based Representation of Emotion Concepts," *Cognition & Emotion* (22:4), Jun, pp 620-632.
- Siemer, M., and Reisenzein, R. 2007. "The Process of Emotion Inference," *Emotion* (7:1), Feb, pp 1-20.
- Simon, H.A. 1997. *Models of Bounded Rationality: Empirically Grounded Economic Reason*. The MIT Press.
- Simons, T. 2002. "Behavioral Integrity: The Perceived Alignment between Managers' Words and Deeds as a Research Focus," *Organization Science* (13:1), Jan-Feb, pp 18-35.

- Skowronski, J.J., and Carlston, D.E. 1987. "Social Judgment and Social Memory - the Role of Cue Diagnosticity in Negativity, Positivity, and Extremity Biases," *Journal of Personality and Social Psychology* (52:4), Apr, pp 689-699.
- Skowronski, J.J., and Carlston, D.E. 1989. "Negativity and Extremity Biases in Impression Formation: A Review of Explanations," *Psychological bulletin*. (105:1), pp 131-142.
- Skowronski, J.J., and Carlston, D.E. 1992. "Caught in the Act - When Impressions Based on Highly Diagnostic Behaviors Are Resistant to Contradiction," *European Journal of Social Psychology* (22:5), Sep-Oct, pp 435-452.
- Smith, C.A., and Ellsworth, P.C. 1985. "Patterns of Cognitive Appraisal in Emotion," *Journal of Personality and Social Psychology* (48:4), pp 813-838.
- Solomon, R.C. 2008. "The Philosophy of Emotions," in: *Handbook of Emotions*, M. Lewis, J.M. Haviland-Jones and L.F. Barrett (eds.). New York: The Guilford Press, pp. 3-16.
- Spool, J.M. 2009. "The Magic Behind Amazon's 2.7 Billion Dollar Question." from <http://www.uie.com/articles/magicbehindamazon/>
- Stefanucci, J.K., and Storbeck, J. 2009. "Don't Look Down: Emotional Arousal Elevates Height Perception," *Journal of Experimental Psychology-General* (138:1), Feb, pp 131-145.
- Straub, D.W. 1989. "Validating Instruments in MIS Research," *MIS Quarterly* (13:2), Jun, pp 147-169.
- Sun, H., and Zhang, P. 2006a. "The Role of Affect in Information Systems Research," in: *Human-Computer Interaction and Management Information Systems: Foundations*, P. Zhang and D.F. Galletta (eds.). Armonk, NY: M.E. Sharpe, pp. 295-329.
- Sun, H.S., and Zhang, P. 2006b. "Causal Relationships between Perceived Enjoyment and Perceived Ease of Use: An Alternative Approach," *Journal of the Association for Information Systems* (7:9), Sep, pp 618-645.

- Tausczik, Y.R., and Pennebaker, J.W. 2010. "The Psychological Meaning of Words: Liwc and Computerized Text Analysis Methods," *Journal of Language and Social Psychology* (29:1), Mar, pp 24-54.
- Tiedens, L.Z. 2001. "The Effect of Anger on the Hostile Inferences of Aggressive and Nonaggressive People: Specific Emotions, Cognitive Processing, and Chronic Accessibility," *Motivation and Emotion* (25:3), Sep, pp 233-251.
- Tiedens, L.Z., and Linton, S. 2001. "Judgment under Emotional Certainty and Uncertainty: The Effects of Specific Emotions on Information Processing," *Journal of Personality and Social Psychology* (81:6), Dec, pp 973-988.
- Toi, M., and Batson, C.D. 1982. "More Evidence That Empathy Is a Source of Altruistic Motivation," *Journal of Personality and Social Psychology* (43:2), pp 281-292.
- Tomlinson, E.C., and Mayer, R.C. 2009. "The Role of Causal Attribution Dimensions in Trust Repair," *Academy of Management Review* (34:1), Jan, pp 85-104.
- Tormala, Z.L., Clarkson, J.J., and Henderson, M.D. 2011. "Does Fast or Slow Evaluation Foster Greater Certainty?," *Personality and Social Psychology Bulletin* (37:3), Mar, pp 422-434.
- Tversky, A., and Kahneman, D. 1974. "Judgment under Uncertainty - Heuristics and Biases," *Science* (185:4157), pp 1124-1131.
- Van Kleef, G.A. 2009. "How Emotions Regulate Social Life: The Emotions as Social Information (Easi) Model," *Current Directions in Psychological Science* (18:3), Jun, pp 184-188.
- Van Kleef, G.A. 2010. "The Emerging View of Emotion as Social Information," *Social and Personality Psychology Compass* (4:5), pp 331-343.
- Veall, M.R., and Zimmermann, K.F. 1996. "Pseudo-R2 Measures for Some Common Limited Dependent Variable Models," *Journal of Economic Surveys* (10:3), pp 241-259.
- Venkatesh, V. 2000. "Determinants of Perceived Ease of Use: Integrating Control, Intrinsic Motivation, and Emotion into the Technology Acceptance Model," *Information Systems Research* (11:4), Dec, pp 342-365.

- Vygotsky, L.S. 1964. "Thought and Language," *Annals of Dyslexia* (14:1), pp 97-98.
- Watson, D., and Tellegen, A. 1985. "Toward a Consensual Structure of Mood," *Psychological Bulletin* (98:2), pp 219-235.
- Weiner, B. 1985. "An Attributional Theory of Achievement-Motivation and Emotion," *Psychological Review* (92:4), pp 548-573.
- Whissell, C. 2009. "Using the Revised Dictionary of Affect in Language to Quantify the Emotional Undertones of Samples of Natural Language," *Psychological Reports* (105:2), Oct, pp 509-521.
- White, G.L., Fishbein, S., and Rutstein, J. 1981. "Passionate Love and the Misattribution of Arousal," *Journal of Personality and Social Psychology* (41:1), pp 56-62.
- Wickens, C.D. 1991. "Processing Resources in Attention," in: *Varieties of Attention*, R. Parasuraman and D.R. Davies (eds.). New York: Academic Press.
- Wyer, R.S. 1973. "Category Ratings as Subjective Expected Values - Implications for Attitude Formation and Change," *Psychological Review* (80:6), pp 446-467.
- Yates, J.F. 1990. *Judgment and Decision Making*. Englewood Cliffs, NJ, US: Prentice-Hall, Inc.
- Yerkes, R.M., and Dodson, J.D. 1908. "The Relation of Strength of Stimulus to Rapidity of Habit-Formation," *Journal of Comparative Neurology and Psychology* (18:5), Nov, pp 459-482.
- Yi, M.Y., and Hwang, Y.J. 2003. "Predicting the Use of Web-Based Information Systems: Self-Efficacy, Enjoyment, Learning Goal Orientation, and the Technology Acceptance Model," *International Journal of Human-Computer Studies* (59:4), Oct, pp 431-449.
- Yiend, J. 2010. "The Effects of Emotion on Attention: A Review of Attentional Processing of Emotional Information," *Cognition & Emotion* (24:1), pp 3-47.
- Yik, M., Russell, J.A., and Steiger, J.H. 2011. "A 12-Point Circumplex Structure of Core Affect," *Emotion* (11:4), Aug, pp 705-731.

Zajonc, R.B. 1980. "Feeling and Thinking - Preferences Need No Inferences," *American Psychologist* (35:2), pp 151-175.

Zajonc, R.B. 1984. "On the Primacy of Affect," *American Psychologist* (39:2), 1984, pp 117-123.

Zeelenberg, R., Wagenmakers, E.J., and Rotteveel, M. 2006. "The Impact of Emotion on Perception," *Psychological Science* (17:4), Apr, pp 287-291.

Zhang, P., and Li, N. 2005. "The Importance of Affective Quality," *Communications of the ACM* (48:9), Sep, pp 105-108.

Zhang, R.C., and Tran, T. 2010. "Helpful or Unhelpful: A Linear Approach for Ranking Product Reviews," *Journal of Electronic Commerce Research* (11:3), pp 220-230.

Zhu, F., and Zhang, X. 2010. "Impact of Online Consumer Reviews on Sales: The Moderating Role of Product and Consumer Characteristics," *Journal of Marketing* (74:2), Mar, pp 133-148.